

Cloudera DataFlow

CDF Security

Date published: 2021-04-06

Date modified: 2024-01-09

CLOUDERA

<https://docs.cloudera.com/>

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

DataFlow security.....	4
DataFlow authorization.....	5
DataFlow authorization process.....	7
Authorize user to manage flow definitions in DataFlow Catalog.....	9
Authorize user to manage lifecycle of a DataFlow environment.....	9
Authorize user to develop flows in a DataFlow environment.....	9
Authorize user to manage deployments in a DataFlow environment.....	10
Authorize user to manage deployments and deploy flows in a DataFlow environment.....	11
Authorize user to view and monitor deployments in a DataFlow environment.....	11
Authorize user to create and manage Projects for a tenant.....	12
Authorize user to create and manage their Projects in a DataFlow environment.....	12
Projects.....	13
Creating a Project.....	14
Editing a Project.....	14
Managing access to a Project.....	15
Adding resources to a Project.....	15
Managing and reassigning resources to a Project.....	16

DataFlow security

As a security administrator, understand the out-of-the box security features as well as the tasks that you need to perform to authorize CDP users.

CDP provides the following security features:

- User authentication
- Role-based user authorization
- Data encryption

You must assign roles to CDP users based on their DataFlow job responsibilities.

User authentication

Users are automatically authenticated through the CDP identity provider. For more information, see *CDP Security Overview*.

Role-based user authorization

DataFlow roles allow you to set user permissions. Through these permissions you can define actions a user or group is allowed to perform and you can also scope the resources a they can perform those actions on.

For more information on the DataFlow roles, see *DataFlow Authorization*.

Data encryption

CDP encrypts data at rest and in motion.

Data at rest

When you import a flow definition, the flow definition is encrypted and stored in the DataFlow Catalog. DataFlow Catalog is a service that enables you to manage flow definitions centrally.



Note:

- Logs and the NiFi FlowFile, Content, and Provenance repositories are stored in your cloud account. Access to the logs and repositories is based on user permissions.
- Before you import a flow definition:
 - Verify that the flow definition will not write secure information to the logs
 - Parameterize sensitive configuration

Data in transit

When you deploy a flow, the encrypted flow definition is transferred to your cloud account where it is decrypted and deployed. During a flow deployment, data can be transferred between servers, systems, applications, and users. Every transfer is a secure and trusted exchange through TLS. Through cryptographic protocols, TLS encrypts and authenticates:

- The flow definition from the DataFlow Catalog to a DataFlow environment.
- The connections from the user's browser to the DataFlow service.
- The communication between the DataFlow service and other services in the CDP Control Plane.
- The communication between the CDP Control Plane services and environment services.
- The communication between the DataFlow service, flows, and shared services within an environment.

Related Concepts

[DataFlow authorization](#)

Related Information

[CDP Security Overview](#)

DataFlow authorization

As a security administrator with the PowerUser role, you must assign one or more DataFlow roles to a user to allow the user to perform tasks in the DataFlow Catalog, the DataFlow environment, and the NiFi cluster.

Review each of the following DataFlow roles to determine which roles a user requires:

DFCatalogViewer

This role allows a user to perform the following tasks:

- View and search flow definitions in the DataFlow Catalog
- View and search ReadyFlows in the ReadyFlow Gallery

DFCatalogAdmin

This role allows a user to centrally manage flow definitions and ReadyFlows.

A user with the DFCatalogAdmin role can perform the following tasks in the DataFlow Catalog:

- View and search flow definitions
- Import flow definitions
- Import new versions of existing flow definitions
- View flow definition details
- Delete flow definitions

A user with the DFCatalogAdmin role can perform the following tasks in the ReadyFlow Gallery:

- View and search ReadyFlows
- Add ReadyFlows to the Catalog

DFAdmin

This role allows a user to manage the lifecycle of a DataFlow environment.



Note: The DFAdmin role does not act as a superset of the DFFlowAdmin and DFFlowUser roles. Users that have only the DFAdmin role cannot view or create flow deployments.

A user with the DFAdmin role can perform the following tasks:

- Enable a CDP environment for DataFlow
- View an environment
- View environment details and DataFlow Settings
- Disable an environment
- Update resource roles
- Grant remote access to an environment
- Revoke remote access to an environment
- Terminate all running deployments when disabling the DataFlow service for an environment.

A user with the DFAdmin and PowerUser roles can perform the following additional tasks:

- Assign the DFAdmin role to other users
- Revoke the DFAdmin role from other users

DFFlowAdmin

This role allows a user to manage flow deployments in a DataFlow environment. In addition, this role together with either the DFCatalogViewer or DFCatalogAdmin role allows the user to deploy flows.

A user with the DFFlowAdmin role alone can perform the following tasks:

- View a deployment on the dashboard
- View deployment details and settings
- View a deployment in NiFi
- Stop a deployment in NiFi
- Suspend a deployment in NiFi
- Terminate a deployment in NiFi
- Change NiFi version

A user with the DFFlowAdmin and PowerUser roles can perform the following additional tasks:

- Assign the DFFlowAdmin and DFFlowUser roles to other users
- Revoke the DFFlowAdmin and DFFlowUser roles from other users

A user with the DFFlowAdmin role and either the DFCatalogViewer or DFCatalogAdmin role can perform the following additional task:

- Deploy flow definitions

DFFlowUser

This role allows a user to view, search, and monitor flow deployments in a DataFlow environment.

A user with the DFFlowUser role can perform the following tasks:

- View a deployment on the dashboard
- View deployment details and settings
- View a deployment in NiFi

DFDeveloper

This role allows a user to view, search, create, and manage drafts in Flow Designer on environment level.

A user with the DFDeveloper role can perform the following tasks:

- Create a new draft
- View all drafts in an environment
- View all draft details and settings in an environment
- Update all drafts in an environment
- Start and end Test Sessions for any draft in an environment
- Delete any draft in an environment

A user with the DFDeveloper and the DFCatalogViewer or DFCatalogAdmin roles can perform the following additional tasks:

- Create drafts from flow definitions in the DataFlow Catalog
- Create drafts from ReadyFlows in the ReadyFlow Gallery
- Publish drafts as flow definitions to the Catalog (DFCatalogAdmin only)

DFProjectsAdmin

This role allows a user to centrally manage Projects.

A user with the DFProjectsAdmin role can perform the following tasks in a tenant:

- View and create Projects.
- Edit users and groups assigned to Projects.
- Delete Projects.

DFProjectCreator

This role allows a user to create and manage Projects on environment level.

A user with the DFProjectCreator role can perform the following tasks:

- Create a Project.
- Edit a Project.
- Manage users and groups assigned to a Project.

For information on configuring CDP user management, see *User Management*.

Related Information

[User Management](#)

DataFlow authorization process

As a security administrator with the CDP PowerUser role, understand the authorization process to allow CDP users to perform tasks.

The following diagram shows the DataFlow authorization process:

Authorize user to manage flow definitions in DataFlow Catalog

To authorize a user to import and manage flow definitions in the DataFlow Catalog and add ReadyFlows to the DataFlow Catalog, you must assign the DFCatalogAdmin role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click User Management.
2. Enter the user's name in the Search field.
3. Select the user from the list.
The **User** page appears.
4. Go to the Roles tab.
5. Select Update Roles.
6. On the **Update Roles** dialog, select the DFCatalogAdmin role along with any other role the user requires.
7. Click Update.

Authorize user to manage lifecycle of a DataFlow environment

To authorize a user to enable, disable, and administer a given CDP environment for DataFlow, assign the DFAdmin role to the user. This role also permits the user to grant and revoke the ability to access the DataFlow Kubernetes API server.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.
- You must have Owner permission for the environment where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click Environments.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click Actions Manage Access .
The **Environment Access** page appears.
4. Find the user and click Update Roles.
5. From the **Update Roles** dialog, select the DFAdmin role and any other role the user requires.
6. Click Update Roles.

Authorize user to develop flows in a DataFlow environment

To authorize a user to create and manage drafts in a DataFlow environment, assign the DFFlowDeveloper role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.
- You must have Owner permission for the environment where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

About this task



Note: To create drafts from flow definitions in the Catalog and ReadyFlows in the ReadyFlow Gallery, users also need the DFFlowViewer or the DFFlowAdmin role. To publish drafts to the Catalog as new flow definitions or versions of existing flow definitions, users need the DFFlowAdmin role.

Procedure

1. From the Cloudera Management Console, click Environments.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click **Actions Manage Access** .
The **Environment Access** page appears.
4. Find the user and click Update Roles.
5. From the **Update Roles** dialog, select the DFFlowDeveloper role.
6. Click Update Roles.
7. Click **Synchronize Users** to propagate the user role changes you made to the environment.



Note: It may take up to five minutes for the user role changes to propagate to the Environment.

Authorize user to manage deployments in a DataFlow environment

To authorize a user to manage flow deployments in a DataFlow environment, assign the DFFlowAdmin role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.
- You must have Owner permission for the environment where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click Environments.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click **Actions Manage Access** .
The **Environment Access** page appears.
4. Find the user and click Update Roles.
5. From the **Update Roles** dialog, select the DFFlowAdmin role.
6. Click Update Roles.

7. Click **Synchronize Users** to propagate the user role changes you made to the environment.



Note: It may take up to five minutes for the user role changes to propagate to the Environment.

Authorize user to manage deployments and deploy flows in a DataFlow environment

To authorize a user to manage flow deployments as well as deploy flows in a DataFlow environment, you must assign the DFFlowAdmin role and either the DFCatalogAdmin or DFCatalogViewer role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click **Environments**.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click **Actions Manage Access** .
The **Environment Access** page appears.
4. Find the user and click **Update Roles**.
5. From the **Update Roles** dialog, select the DFFlowAdmin role.
6. Click **Update Roles**.
7. Click **Synchronize Users** to propagate the user role changes you made to the environment.



Note: It may take up to five minutes for the user role changes to propagate to the Environment.

8. From the Cloudera Management Console, click **User Management**.
9. Enter the user's name in the **Search** field.
10. Select the user from the list.
The **User** page appears.
11. Go to the **Roles** tab.
12. Select **Update Roles**.
13. On the **Update Roles** dialog, assign one of the following DFCatalog roles to the user:
 - If the user needs to deploy flow definitions in the environment, assign the DFCatalogViewer role.
 - If the user needs to manage flow definitions in the DataFlow Catalog and deploy flow definitions in the environment, assign the DFCatalogAdmin role.
14. Click **Update**.

Authorize user to view and monitor deployments in a DataFlow environment

To authorize a user to view, search, and monitor deployments in a DataFlow environment, assign the DFFlowUser role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.
- You must have Owner permission for the environment where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click Environments.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click Actions Manage Access .
The **Environment Access** page appears.
4. Find the user and click Update Roles.
5. From the **Update Roles** dialog, select the DFFlowUser role and any other role the user requires.
6. Click Update Roles.
7. Click Synchronize Users to propagate the user role changes you made to the environment.



Note: It may take up to five minutes for the user role changes to propagate to the Environment.

Authorize user to create and manage Projects for a tenant

To Authorize a user to create and delete Projects, edit Users and Groups assigned to Projects within a tenant, you must assign the DFProjectsAdmin role to the user.

Before you begin

You must have the CDP PowerUser role.

Procedure

1. From the Cloudera Management Console, click User Management.
2. Enter the user's name in the Search field.
3. Select the user from the list.
The **User** page appears.
4. Go to the Roles tab.
5. Select Update Roles.
6. On the **Update Roles** dialog, select the DFProjectsAdmin role along with any other role the user requires.
7. Click Update.

Authorize user to create and manage their Projects in a DataFlow environment

To authorize a user to create and manage their own Projects, edit Users and Groups assigned to Projects within an environment, you must assign the DFProjectCreator role to the user.

Before you begin

- You must have PowerUser permission for the account where you want to perform this task.

- You must have Owner permission for the environment where you want to perform this task.
- You must have IAMViewer or IAMUser permission for the account where you want to perform this task.

Procedure

1. From the Cloudera Management Console, click Environments.
The **Environment List** page appears.
2. Search for and select the environment you want the user to have rights to.
The **Environment** page appears.
3. Click **Actions Manage Access** .
The **Environment Access** page appears.
4. Find the user and click **Update Roles**.
5. From the **Update Roles** dialog, select the DFProjectCreator role.
6. Click **Update Roles**.
7. Click **Synchronize Users** to propagate the user role changes you made to the environment.



Note: It may take up to five minutes for the user role changes to propagate to the Environment.

Projects

Projects provide fine grained role-based access control over resources within an environment.

Projects are a layer of role-based access control in CDF. You may consider them as logical containers that you can use to govern access to resources within an environment.

Resources (flow deployments, flow drafts, Inbound Connections, and custom NARs) can have one of two states within an environment:

- Unassigned - the resource is not assigned to a particular Project, it is freely accessible for every user or group with the appropriate user role for that environment.
- Assigned to a Project - the resource is only accessible to users or groups with access to that particular Project. This assignment is exclusive, you cannot assign the same resource to more than one Project at a time.

The roles associated with projects are additive, they work in conjunction with user permissions controlling the types of actions a user or group is allowed to perform in a CDP environment.

Projects introduces the following user roles:

DFProjectsAdmin

This role enables users to manage all Projects in all environments within an tenant.

DFProjectCreator

This role enables users to create Projects in an Environment. When creating a Project, they automatically become the DFProjectAdmin in that Project.

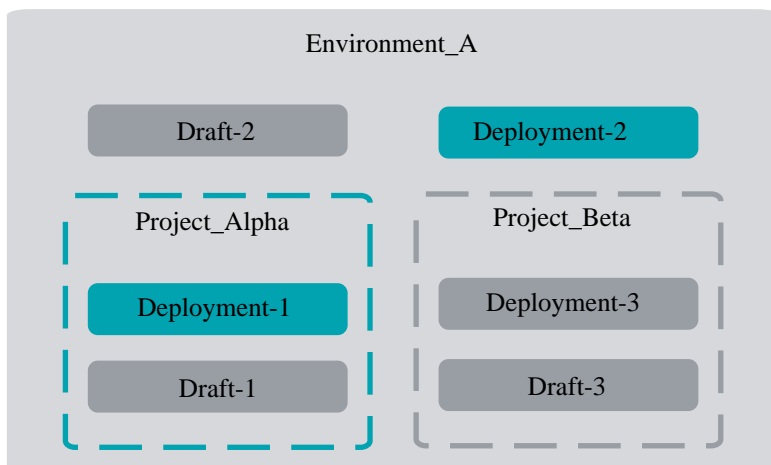
DFProjectAdmin

This role is automatically assigned to users with DFProjectCreator role upon creating a Project. DFProjectAdmins can add/remove users and groups, changing user and group roles (DFProjectAdmin or DFProjectMember), modify the Project name and description, and delete the Project.

DFProjectMember

This role enables users to access resources assigned to a Project and perform actions on them that are allowed by their other user roles.

For example, users with DFFlowUser role in 'Environment_A', and DFProjectMember role in 'Project_Alpha' are able to view flow deployments 'Deployment-1' that is assigned to 'Project_Alpha' and 'Deployment-2' that is currently unassigned. They cannot view 'Deployment-3' that is assigned to Project_Beta, just as they do not see other resource types (in this example draft flows 'Draft-1', 'Draft-2', and 'Draft-3') regardless of their assignment status. That is, the DFFlowUser role defines the types of actions they can perform in the environment, and the 'DFProjectMember' role defines the resources within that environment they can perform those actions on.



Creating a Project

You can create a Project to limit the visibility of drafts, deployments, Inbound Connections, and custom NAR Configurations within an Environment.

Before you begin

- Make sure that you have DFProjectCreator permission to perform this task. For information on account and resource roles, see DataFlow authorization.
- DataFlow must be enabled for the Environment.

Procedure

1. Go to DataFlow Projects .
2. Click Create Project.
3. Select Target Workspace.
4. Provide Project Name.



Note: Project names need to be unique within a Workspace. The Deployment wizard indicates whether a name is valid by displaying a green check below the Project Name text box.

5. Provide a Project Description.
6. Click Create.



Note: After creating a Project, it needs to be synchronized before resources can be assigned. This may take a few minutes and during the synchronization, the new Project is not available in the Workspace Resources view yet.


Editing a Project

You can change the name and the description of Projects where you have DFProjectAdmin permission.

Before you begin

- Make sure that you have DFProjectAdmin permission for the project you want to edit to perform this task.

Procedure

1. Go to DataFlow Projects .
2. Select the Project you want to edit and click  More Edit Project .
3. Modify the Project name and/or Project description.
4. Click Apply Changes.


Managing access to a Project

You can grant and revoke access to Projects where you have DFProjectAdmin permission.

Before you begin

- Make sure that you have DFProjectAdmin permission for the project you want to manage to perform this task.
- Make sure that you have IAMViewer or IAMUser permission for the tenant where you want to perform this task. Without one of those, users and groups will not be visible to you.

Procedure

1. Go to DataFlow Projects .
2. Select the project where you want to manage access rights and click  More Manage Access .

To grant roles:

3. Start typing the name of the user or group you want to add and select them from the list.
4. Select the Resource Roles you want to grant.
5. Click Update Roles.
6. Click Synchronize Users.

To revoke roles:

7. Click Update Roles next to the name of the user or group for which you want to revoke roles.
8. Unselect the Resource Roles you want to revoke.
9. Click Update Roles.
10. Click Synchronize Users.

Adding resources to a Project


You can add resources to Projects where you have DFProjectMember permission.

Before you begin


- Make sure that you have DFProjectMember permission for the project where you want to add resources to perform this task.

Procedure

1. Go to DataFlow Projects .

2. Select the project where you want to add resources and click  More View Workspace Manage Project Resources .

The **Workspace Resources** view shows you all resources across Projects that you have access to.

3. Select the type of resource you want to reassign using the resource type tabs.
4. Select the resource you want to reassign and click  More Reassign .



Tip: Select several resources using the checkboxes and click the Reassign button for bulk resource reassignment.

5. In the **Reassign Resource** pop-up select an available Project or select Unassigned.
6. Click Reassign.


Managing and reassigning resources to a Project


You can manage and reassign resources to Projects where you have DFProjectMember permission.

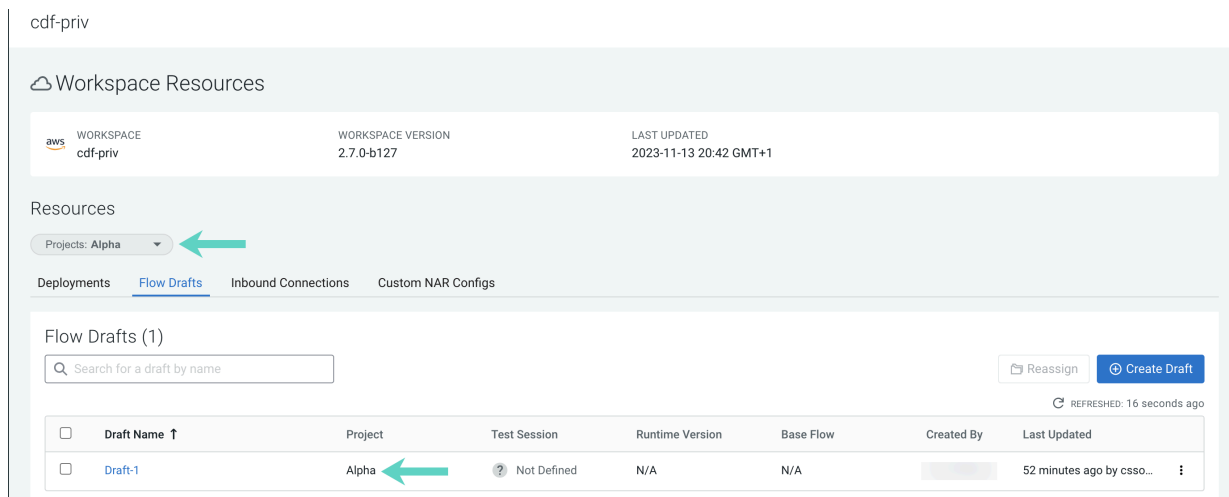
Before you begin

- You must have DFProjectMember permission for the Project where you want to perform this task.


Procedure

1. Go to DataFlow Projects .
2. Select the project where you want to manage resources and click  More View Project Resources - Manage & Assign .

The  **Workspace Resources** view shows you resources filtered for the Project you selected.




The screenshot displays the 'Workspace Resources' view for a workspace named 'cdf-priv'. It shows workspace details like version '2.7.0-b127' and last updated time '2023-11-13 20:42 GMT+1'. Below, there are tabs for 'Resources', 'Deployments', 'Flow Drafts', 'Inbound Connections', and 'Custom NAR Configs'. The 'Flow Drafts' tab is selected, showing a search bar and a table of drafts. A red arrow points to the 'Projects: Alpha' dropdown menu, and another red arrow points to the 'Alpha' project in the table.

Draft Name ↑	Project	Test Session	Runtime Version	Base Flow	Created By	Last Updated
<input type="checkbox"/> Draft-1	Alpha	 Not Defined	N/A	N/A		52 minutes ago by cspo...



Note: After creating a Project, it needs to be synchronized before resources can be assigned. This may take a few minutes and during the synchronization, the new Project is not available in the Workspace Resources view yet.

3. Select the type of resource you want to reassign using the resource type tabs.

4. Select the resource you want to reassign and click  More Reassign .



Tip: Select several resources using the checkboxes and click the Reassign button for bulk resource reassignment.

5. In the **Reassign Resource** pop-up select an available Project or select Unassigned.
6. Click Reassign.