Cloudera DataFlow

Managing DataFlow in an Environment

Date published: 2021-04-06
Date modified: 2024-01-09

https://docs.cloudera.com/

Legal Notice

Contents

Managing DataFlow in an environment

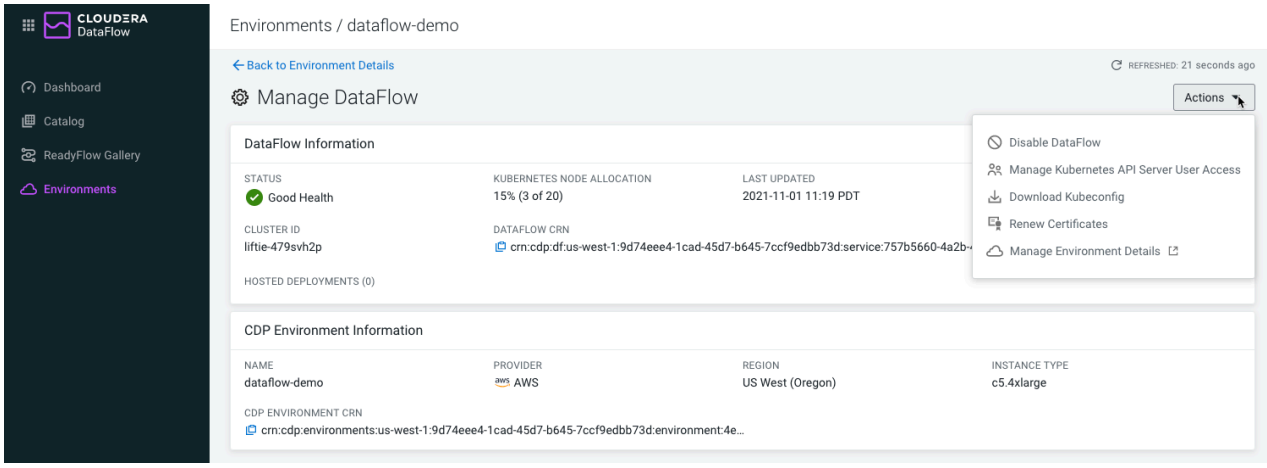
You can use the DataFlow Manager page to manage and monitor your DataFlow environment.

The Actions drop-down menu in the DataFlow Manager page allows you to choose between the following options to manage DataFlow in an environment:

- Disable DataFlow for the environment
- Reset DataFlow for the environment
- Manage user acces for the Kubernetes API Server
- Download the Kubeconfig file
- Renew certificates
- Manage the environment details in Management Console

Apart from the information on your DataFlow Environment, the DataFlow Manager page also displays the capacity, networking, Kubernetes API Server endpoint access and tags of your environment under DataFlow Settings. You can edit the capacity settings of the environment, update the IP address ranges that are allowed to access the Kubernetes API Server and Load Balancer, and review the tags associated with the DataFlow environment under DataFlow Settings.

Click Manage DataFlow, from the Environment Details pane to perform some actions on your environment. The DataFlow Manager page appears.



You can also go back to the environment details by clicking Back to Environment Details.

Disabling DataFlow for an environment

Disabling DataFlow for an environment terminates the cloud infrastructure that was created as part of the enablement process.

About this task

When you disable DataFlow for an environment, you can specify whether to preserve your environment event history. Preserving the event history retains your environment event history and allows you to view past events even after the environment has been disabled. Regardless of whether you choose to preserve the event history, you can enable DataFlow for an environment again after a successful disablement operation.

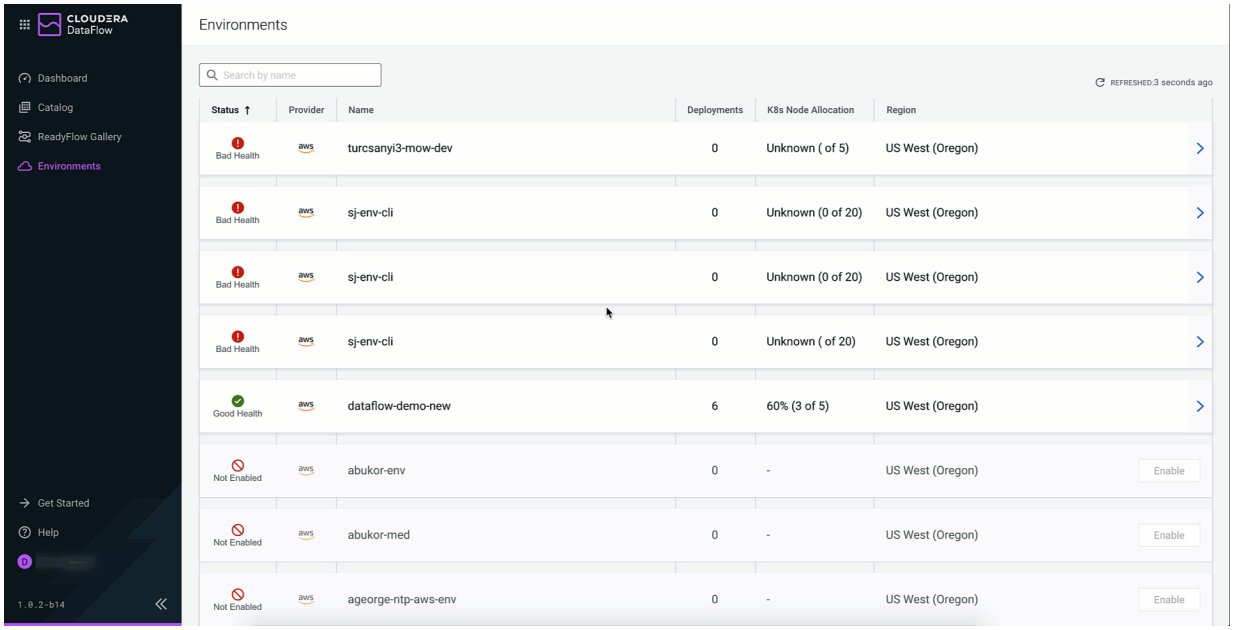
Steps

For UI

1. In DataFlow, from the Environments page, select the environment you want to disable.
2. Click Manage DataFlow from the **Environment Details** pane.

   You are redirected to the Manage DataFlowpage.
3. From the Actions menu, select Disable DataFlow.
4. Specify whether you want to Preserve event history.
5. Enter the environment name to confirm.
6. Select Disable to initiate the disablement process.

Example



For CLI

Before you begin

- You have installed CDP CLI.
- Run cdp df list-services to get the service-crn.

1. To disable DataFlow for an environment, enter:

```
cdp df disable-service
--service-crn [***SERVICE_CRN***]
[--persist] [--no-persist]
[--terminate-deployments] [--no-terminate-deployments]
                [help]
```

Where:

- service-crn – Provides the value you identified when you run cdp df list-services.
- [--persist] [--no-persist] – Select one to specify whether you want to preserve environment history.
- [--terminate-deployments] [--no-terminate-deployments] – Specifies whether you want to gracefully terminate deployments associated with this environment. Regardless of this setting all associated deployments will be terminated when you disable DataFlow

Result

When you successfully disable DataFlow for an environment, your result will be similar to:

```
{
  "status": {
    "state": "DISABLING",
    "message": "Disabling DataFlow",
    "detailedState": "TERMINATING_DEPLOYMENTS"
  }
}
```

Next steps

Disabling DataFlow for an environment can take up to 30 minutes.

Related Information

Clearing the DataFlow environment event history

Resetting your environment

Managing remote access

Downloading kubeconfig

Clearing the DataFlow environment Event History

About this task

When disabling DataFlow for an environment, you can choose to preserve the Event History for the specific environment. This allows you to review past events even after DataFlow has been disabled for an environment. When the preserved events are no longer relevant, you can delete them by using the Clear Event History action.

Note:

The Clear event history action is only available for disabled DataFlow environments with a preserved event history.

Before you begin

- You have the DFAdmin user role for the environment for which you want to clear the event history.

Procedure

1. Select the environment for which you want to clear the event history.
2. Click Manage DataFlow from the **Environment Details** pane.
   You are redirected to the **Manage DataFlow** page.
3. From the Actions menu, select Clear event history.
4. Select Clear Event History to confirm deleting all event-related information and past alert conditions.

Results

After successfully clearing the event history, you are no longer able to view the environment details by clicking on it. You can enable DF in the environment again by using the Enable button on the environment row.

Related Information

Disabling DataFlow for an environment

Resetting your environment

Managing remote access

Downloading kubeconfig

Resetting your environment

When disabling DataFlow for a specific environment fails, you can use the Reset Environment action to reset an environment state for DataFlow

Before you begin

- You have the DFAdmin user role for the environment you want to reset.

Procedure

1. In DataFlow, from the Environment page, select the environment you want to reset.
2. Click Manage DataFlow from the **Environment Details** pane.
   You are redirected to the **Manage DataFlow** page.
3. From the Actions menu, select Reset Environment.
4. Click Reset in the confirmation dialog to proceed.

Results

Resetting an environment clears DataFlow state without impacting the associated CDP environment and any of its components including Data Hubs, Data Lakes, and FreeIPA. If the associated CDP environment is still healthy, resetting allows you to enable it again for DataFlow.

Note:

Resetting an environment does not delete associated cloud resources which were created during its enablement process. Manual steps may be necessary to address these orphaned resources in your cloud account.

Related Information

Disabling DataFlow for an environment
Clearing the DataFlow environment event history
Managing remote access
Downloading kubeconfig

Manage Kubernetes API Server user access

Giving users remote access to CDF environments allows authorized users to use kubectl to manage and troubleshoot Kubernetes clusters using the Kubernetes API. To do this, use the Actions menu from the Environments page.

About this task

The API server of the Kubernetes cluster which is created when enabling a CDP environment for DataFlow is secured using authentication and role based access control. By default no one is allowed to connect to the Kubernetes API server. You can grant users access to the Kubernetes API server by adding their AWS ARN to the list of Authorized Users so they can communicate with the cluster using Kubernetes management tools such as `kubectl`.

Before you begin

- You have the DFAdmin user role.

- You have a cloud user ID. For AWS this is an ARN and looks similar to:

```
arn:aws:iam::{AWSaccountID}:role/{IAMRoleName}
```

See the *AWS documentation* for more information.

Procedure

1. In DataFlow, from the Environments page, click the Environment for which you want to add or remove user access.
2. Click Manage DataFlow.
3. From the Actions menu, click Manage Kubernetes API Server User Access.
4. Provide the Cloud User ID you want to authorize.
   - To add more than one user, add Cloud User IDs one by one.
   - To remove a user, click the remove icon for the particular row.

What to do next

Download the kubeconfig file and share it with authorized users so they can connect to the cluster using their preferred Kubernetes management tools

Related Information

Amazon EKS IAM roles
Disabling DataFlow for an environment
Clearing the DataFlow environment event history
Resetting your environment
Downloading kubeconfig

Renewing certificates

Certificates for accessing DataFlow have a 90 day lifespan. They are automatically renewed after 60 days. Should you need to manually renew your certificates you can use the Actions menu to do so.

Procedure

1. From the Environments page, click Actions.
2. Select Renew Certificates from the drop-down menu.
3. Confirm by clicking Renew Certificates.

Note:

Renewing your certificates assigns new certificates for accessing DataFlow. The old certificates are not revoked.

Downloading kubeconfig

You can download the kubeconfig file so that you can use the `kubectl` management tool to manage and troubleshoot your CDF Kubernetes cluster.

About this task

After granting users access to the Kubernetes API server, you can download the Kubeconfig for a Kubernetes cluster so they can communicate with it using Kubernetes management tools such as `kubectl`.

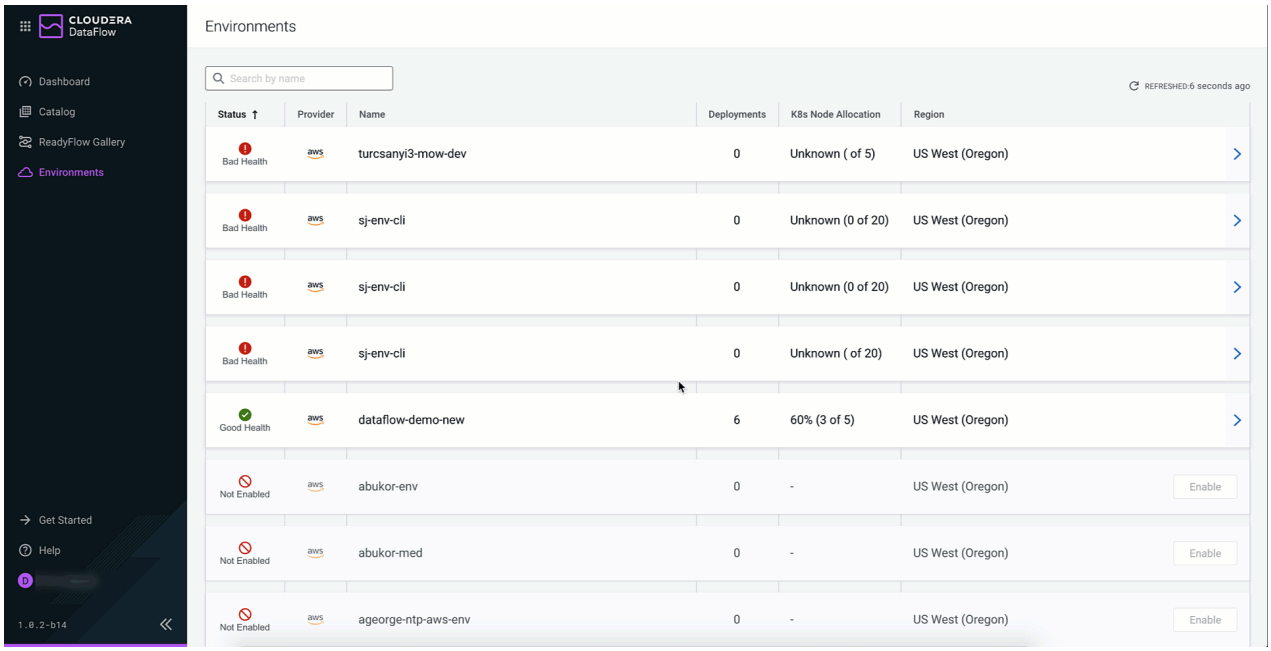
Before you begin

- You have the DFAdmin user role for the environment.

Procedure

1. In DataFlow, from the Environments page, click the environment for which you want to download the kubeconfig file.
2. Click Manage DataFlow from the **Environment Details** pane.
   You are redirected to the **Manage DataFlow** page.
3. From the Actions menu, click Download Kubeconfig.
4. Share the kubeconfig file with authorized users.

Example



Related Information

Disabling DataFlow for an environment
Clearing the DataFlow environment event history
Resetting your environment
Managing remote access

Updating node images in a DataFlow service

Learn about adopting a new node image for your DataFlow service.

About this task

This action updates the cloud provider node images of your environment. During the update you will not be able to create or manage deployments, drafts, or test sessions in the DataFlow service.

Before you begin

- You have the DFAdmin user role for the environment where you want to update the node images.

Procedure

1. In DataFlow, from the Environments page, select the Environment where you want to update the node images.
2. From the Actions menu select Update Node Images.
3. Click Update.

   The service status changes to Updating. Wait until the status returns to Good Health before initiating any other action.