

Managing DataFlow in an Environment

Date published: 2021-04-06

Date modified: 2024-06-03

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

Managing in an environment.....	4
Disabling for an environment.....	4
Clearing the environment Event History.....	6
Resetting your environment.....	7
Managing Kubernetes API Server user access.....	7
Downloading kubeconfig.....	8
Renewing certificates.....	9
Updating Kubernetes node images in a service.....	9
Configuring access for NiFi metrics scraping.....	10

Managing in an environment

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

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

- Disable for the environment
- Reset for the environment
- Manage user access for the Kubernetes API Server
- Download the Kubeconfig file
- Renew certificates
- Manage the environment details in
- Configure NiFi metrics access

Apart from the information on your 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 environment under DataFlow Settings.

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

The screenshot shows the Cloudera DataFlow Manager interface. On the left is a dark sidebar with navigation links: Dashboard, Catalog, ReadyFlow Gallery, and Environments (highlighted). The main content area is titled 'Environments / dataflow-demo' and 'Manage DataFlow'. It includes a 'Back to Environment Details' link and a refresh indicator ('REFRESHED: 21 seconds ago'). An 'Actions' dropdown menu is open, showing options: Disable DataFlow, Manage Kubernetes API Server User Access, Download Kubeconfig, Renew Certificates, and Manage Environment Details. The 'DataFlow Information' section displays: STATUS (Good Health), KUBERNETES NODE ALLOCATION (15% (3 of 20)), LAST UPDATED (2021-11-01 11:19 PDT), CLUSTER ID (lif1e-479svh2p), DATAFLOW CRN (cm.cdp.df.us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb73d:service:757b5660-4a2b...), and HOSTED DEPLOYMENTS (0). The 'CDP Environment Information' section displays: NAME (dataflow-demo), PROVIDER (AWS), REGION (US West (Oregon)), INSTANCE TYPE (c5.4xlarge), and CDP ENVIRONMENT CRN (cm.cdp.environments:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb73d:environment:4e...).

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

Disabling for an environment

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

About this task

When you disable 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 for an environment again after a successful disablement operation.

Steps

For UI

1. In , 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 DataFlow page.

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

Status ↑	Provider	Name	Deployments	K8s Node Allocation	Region
Bad Health	aws	turcsanyi3-mow-dev	0	Unknown (of 5)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (0 of 20)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (0 of 20)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (of 20)	US West (Oregon)
Good Health	aws	dataflow-demo-new	6	60% (3 of 5)	US West (Oregon)
Not Enabled	aws	abukor-env	0	-	US West (Oregon)
Not Enabled	aws	abukor-med	0	-	US West (Oregon)
Not Enabled	aws	ageorge-ntp-aws-env	0	-	US West (Oregon)

For CLI

Before you begin

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

1. To disable 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

Result

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

```
{
```

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

Next steps

Disabling for an environment can take up to 30 minutes.

Related Information

[Clearing the environment event history](#)

[Resetting your environment](#)

[Managing remote access](#)

[Downloading kubeconfig](#)

Clearing the environment Event History

About this task

When disabling for an environment, you can choose to preserve the Event History for the specific environment. This allows you to review past events even after 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 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 for an environment](#)

[Resetting your environment](#)

[Managing remote access](#)

[Downloading kubeconfig](#)

Resetting your environment

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

Before you begin

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

Procedure

1. In `,` 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 `state` without impacting the associated `environment` and any of its components including Data Hubs, Data Lakes, and FreeIPA. If the associated `environment` is still healthy, resetting allows you to enable it again for `.`



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 `for` an environment](#)

[Clearing the `event` history](#)

[Managing remote access](#)

[Downloading kubeconfig](#)

Managing Kubernetes API Server user access

Giving users remote access to `-enabled` 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 `environment` for `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 , 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 for an environment](#)

[Clearing the environment event history](#)

[Resetting your environment](#)

[Downloading kubeconfig](#)

Downloading kubeconfig

You can download the kubeconfig file so that you can use the `kubectl` management tool to manage and troubleshoot your 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 , 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

Status ↑	Provider	Name	Deployments	K8s Node Allocation	Region
Bad Health	aws	turcsanyi3-mow-dev	0	Unknown (of 5)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (0 of 20)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (0 of 20)	US West (Oregon)
Bad Health	aws	sj-env-cli	0	Unknown (of 20)	US West (Oregon)
Good Health	aws	dataflow-demo-new	6	60% (3 of 5)	US West (Oregon)
Not Enabled	aws	abukor-env	0	-	US West (Oregon)
Not Enabled	aws	abukor-med	0	-	US West (Oregon)
Not Enabled	aws	ageorge-ntp-aws-env	0	-	US West (Oregon)

Related Information

[Disabling for an environment](#)

[Clearing the environment event history](#)

[Resetting your environment](#)

[Managing remote access](#)

Renewing certificates

Certificates for accessing 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. The old certificates are not revoked.

Updating Kubernetes node images in a service

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

About this task

This action updates the images to the latest available version on the Kubernetes (K8s) nodes that form the underlying cluster in your service.



Note: This is primarily a troubleshooting option, you do not need to perform this as a routine maintenance task.





Important: During the update you are not able to create or manage deployments, drafts, or test sessions in the service.

Before you begin

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

Procedure

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

- If there is a newer node image than the one already installed, the service status changes to  Updating.

Wait until the status returns to  Good Health before initiating any other action.

- If there is no newer version of the node image available, you get the message:
'No new node image is available, nothing to update.'

Configuring access for NiFi metrics scraping

You can configure an external Prometheus service to scrape NiFi metrics for deployments. To do that, you need to generate a password and add a job for each deployment to your Prometheus configuration.

Before you begin

- You have the DFAdmin user role for the environment where you want to configure access for NiFi metrics scraping.

About this task




Tip: If you want to bulk add deployments to your Prometheus configuration, create a CLI script that collects deployment names and adds the required jobs with the generated password.



Note:

If you need the Endpoint URL of a deployment, go to Deployments [***DEPLOYMENT NAME***] Actions Manage Deployment NiFi Configuration and copy the **NIFI METRICS ENDPOINT URL**.

Procedure

- In , from the Environments page, select the Environment where you want to configure NiFi metrics scraping.
- From the Actions menu select  Access NiFi Metrics.

- Depending on whether you are setting up access for the first time or updating an existing one, select **Initial configuration** or **Manage existing**.

For Initial configuration

- Click Generate Credentials and Enable Access.

Copy the generated password. The username is nifi-metrics for all jobs, do not change it.



Note: You will not be able to access the generated credentials after closing the dialog box.

- Create a new job for each deployment where you want to perform metrics scraping and add it to your Prometheus configuration. Depending on your use case, either append it to an existing configuration or you can create a new one. You can use the provided **Sample Prometheus scrape configuration**.

Figure 1: Sample metrics configuration code snippet

```
scrape_configs:
- job_name: 'nifi-metrics-[***DEPLOYMENT-NAME***]'
  scrape_interval: 15s

  scheme: https
  honor_labels: true
  metrics_path: /dfx-[***DEPLOYMENT-NAME***]-ns/federate

  basic_auth:
    # Use 'nifi-metrics' as the username for all jobs.
    [ username: nifi-metrics ]
    [ password: [***GENERATED PASSWORD***] ]

  params:
    'match[]':
      # This parameter is mandatory, because Cloudera's Prometheus
      instance also scrapes cadvisor and Prometheus itself.
      - '{job="dfx-nifi-web"}'

  static_configs:
    - targets: ['https://dfx.qbllchii.xcu2-8y8x.dev.cldr.work']
```

You need to replace

- [***DEPLOYMENT-NAME***] with the encoded deployment name. (For example 'Some DataFlow Deployment' is encoded as 'some-dataflow-deployment')
- [***GENERATED PASSWORD***] with the generated password.

Depending on your Prometheus setup, you may need to make further additions to the job definition.



Tip:

To find a deployment-specific YAML snippet sample where you only need to substitute the generated password, go to Deployments [***DEPLOYMENT NAME***] Actions Manage Deployment NiFi Configuration and copy the **SAMPLE PROMETHEUS SCRAPE CONFIGURATION**.

- Add the newly created or updated configuration file to your Prometheus service.

For Manage existing

- To turn on or off NiFi metrics scraping for all flow deployments in an environment, toggle the Access NiFi Metrics switch. Turning metrics scraping off stops exposing the Prometheus endpoint.

- If you need to reset the generated credentials, click Regenerate Credentials. Copy the generated password. The username is nifi-metrics for all jobs, do not change it. You will not be able to access the generated credentials after closing the dialog box.



Note: Keep in mind that regenerating the credentials affects all flows in the given environment.

Do not forget to update your Prometheus configuration with the regenerated credentials.

- To add new jobs or remove existing ones, modify the Prometheus configuration file.
- a. Create a new job for each deployment where you want to perform metrics scraping and add it to your Prometheus configuration. Add it to the existing configuration. You can use the provided **Sample Prometheus scrape configuration**.

Figure 2: Sample metrics configuration code snippet

```
scrape_configs:
- job_name: 'nifi-metrics-[***DEPLOYMENT-NAME***]'
  scrape_interval: 15s

  scheme: https
  honor_labels: true
  metrics_path: /dfx-[***DEPLOYMENT-NAME***]-ns/federate

  basic_auth:
    # Use 'nifi-metrics' as the username for all jobs.
    [ username: nifi-metrics ]
    [ password: [***GENERATED PASSWORD***] ]

  params:
    'match[]':
      # This parameter is mandatory, because Cloudera's Prometheus
      # instance also scrapes cadvisor and Prometheus itself.
      - '{job="dfx-nifi-web"}'

  static_configs:
    - targets: ['https://dfx.qbllchii.xcu2-8y8x.dev.cldr.work']
```

You need to replace

- [***DEPLOYMENT-NAME***] with the encoded deployment name. (For example 'Some DataFlow Deployment' is encoded as 'some-dataflow-deployment')
- [***GENERATED PASSWORD***] with the generated password.

Depending on your Prometheus setup, you may need to make further additions to the job definition.



Tip:

To find a deployment-specific YAML snippet sample where you only need to substitute the generated password, go to Deployments [***DEPLOYMENT NAME***] Actions Manage Deployment NiFi Configuration and copy the **SAMPLE PROMETHEUS SCRAPE CONFIGURATION**.

- b. Add the updated configuration file to your Prometheus service.