Cloudera DataFlow

Managing Flow Deployments

Date published: 2021-04-06
Date modified: 2023-11-01

https://docs.cloudera.com/

Legal Notice

Contents

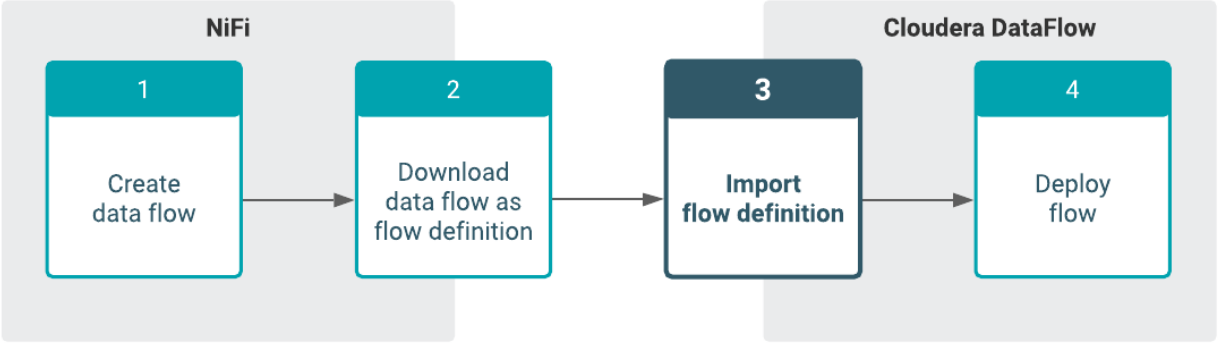
Importing flow definitions from NiFi.....4

Accessing Deployment Manager.....4

Viewing data flow in NiFi.....5
Starting a flow.....6
Stopping a flow.....6
Changing NiFi runtime version.....7
Downloading NiFi application log.....8
Suspending a deployment.....8
Resuming a deployment.....9
Export deployment configuration.....10
Terminating a deployment.....10
Editing a deployment.....11

Importing flow definitions from NiFi

If you want to use a NiFi flow in Cloudera DataFlow, you must import it as a flow definition. When imported, the flow definition is added to the flow Catalog.



Note: For information on publishing a draft from Flow Designer, see *Managing Drafts*.

Accessing Deployment Manager

You can use the Deployment Manager page to manage your flow deployment lifecycle.

About this task

The Actions drop-down menu in the Deployment Manager page allows you to view your flow deployment in NiFi for troubleshooting purposes. It also allows you to change NiFi runtime version. You can suspend a flow deployment. Suspending a flow deployment means that the flow deployment stops processing data but cloud resources remain allocated for the flow deployment. You can also terminate your flow deployment. Terminating a flow deployment deletes the associated NiFi resources.

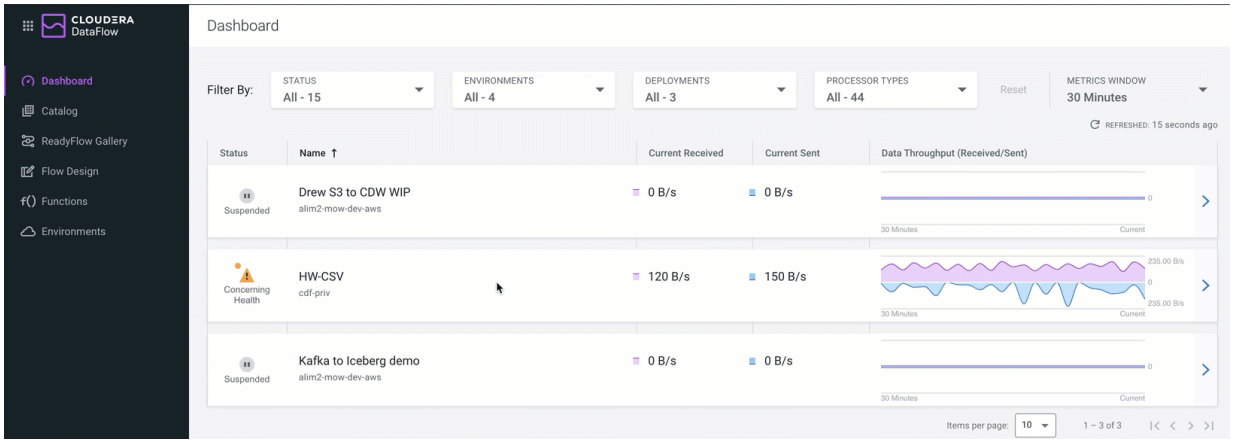
Apart from the details of your flow deployment, the Deployment Manager page also displays KPIs and alerts, sizing and scaling, and parameters details of your flow deployment under Deployment Settings. You can edit the KPIs, alerts, parameters, size and scale details of your flow deployment under Deployment Settings.

To access the Deployment Manager page:

Procedure

1. Open Cloudera DataFlow by clicking the DataFlow tile in the CDP sidebar.
2. Select the flow deployment you want to manage, expanding the Deployment Details pane.

3. Click Manage Deployments.



The Deployment Manager page appears.

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

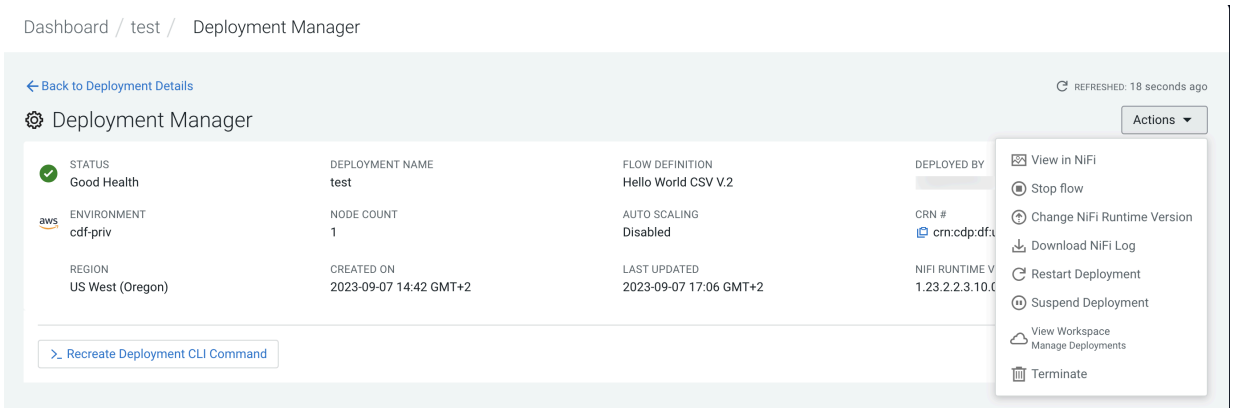
When you access the NiFi cluster, the ability to view or edit the flow is based on your DataFlow authorizations. The DFFlowUser role has read-only privileges. The DFFlowAdmin role has full privileges.

Before you begin

You must have deployed a data flow in Cloudera DataFlow.

Procedure

1. Click Actions in the **Deployment Manager** page.



2. Click 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*.

Related Information

Downloading a flow definition from NiFi
Importing a flow definition to Cloudera DataFlow
Deploy a flow

Starting a flow

You can start a stopped flow a DataFlow deployment.

About this task

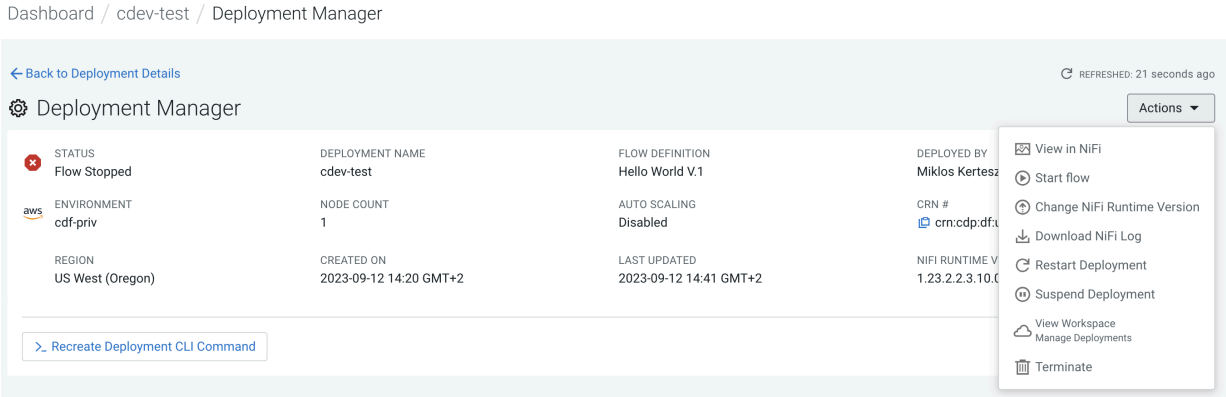
Starting a flow deployment starts all processors of a DataFlow deployment.

Before you begin

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

Procedure

1. Click Actions.



2. Click Start flow.
   The Start [Deployment Name] pop-up appears.
3. Click Start Flow.

Stopping a flow

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

About this task

Stopping a flow results in the following:

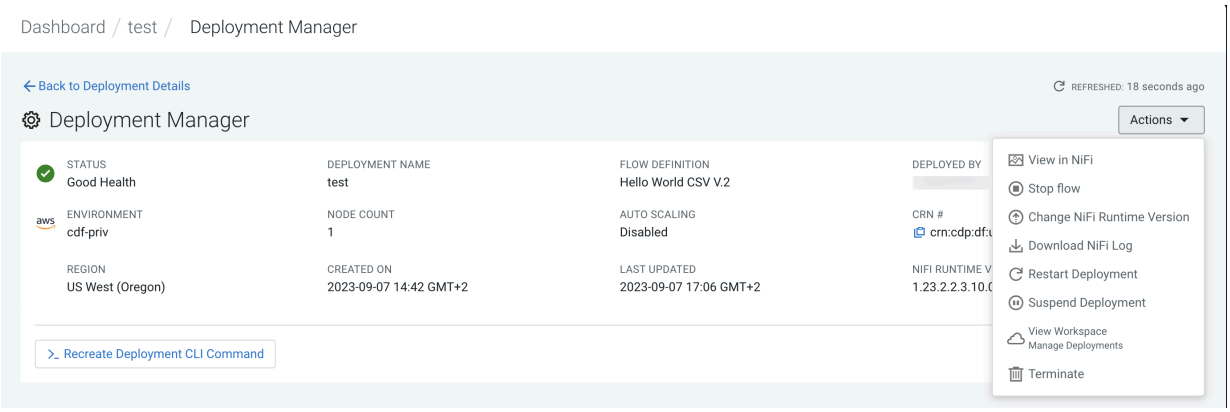
- All processors are stopped and no data processing happens within the NiFi flow.
- KPI alerts are stopped. Your KPI alerts are activated again when the flow is restarted.
- Any active KPI alerts are resolved.
- All underlying cloud resources remain allocated for the DataFlow deployment.
- You can modify 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. Click Actions in the **Deployment Manager** page.



2. Click Stop flow.

   The Stop [Deployment Name] page appears.

3. Click Stop Flow to stop the flow deployment.

Changing NiFi runtime version

You can change the NiFi runtime version for your flow deployment. Generally, you change NiFi runtime version to pick up hotfixes and the latest NiFi version. However, when you create a flow deployment you can also pick the NiFi version to align with your flow certification needs. Cloudera recommends that you always use the latest NiFi version if possible.

About this task

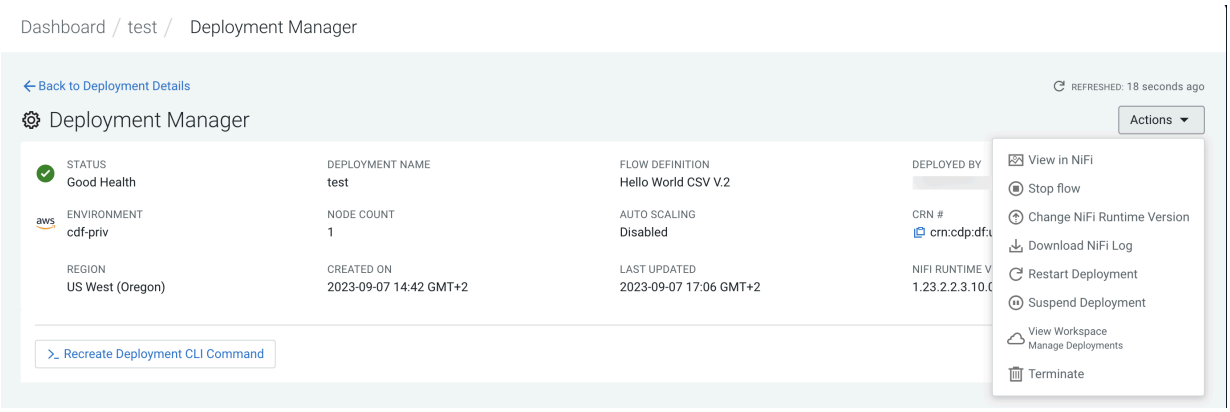
If you opt to change the NiFi runtime version, the process is carried out node by node. If your flow deployment has multiple nodes, the flow deployment continues to run. If the flow deployment has only one node, it stops for a short period of time.

Before you begin

You must have deployed a flow definition in Cloudera DataFlow and you must have the DFFlowAdmin role.

Procedure

1. Click Actions in the **Deployment Manager** page.



2. Click Change NiFi Runtime Version.

   The Change NiFi Runtime Version page appears. It shows the current NiFi runtime version and a list of new version numbers.

3. Select a new version number to update from the New Version dropdown, and click Update.

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. On the Dashboard select the deployment for which you want to download the NiFi application log.
2. Select Actions Manage Deployment .
   You are redirected to the **Deployment Manager** page.
3. Select Actions Download NiFi Log .

Results

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

Related Information

Diagnostic bundle collection

Suspending a deployment

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

About this task

Suspending a DataFlow deployment results in the following:

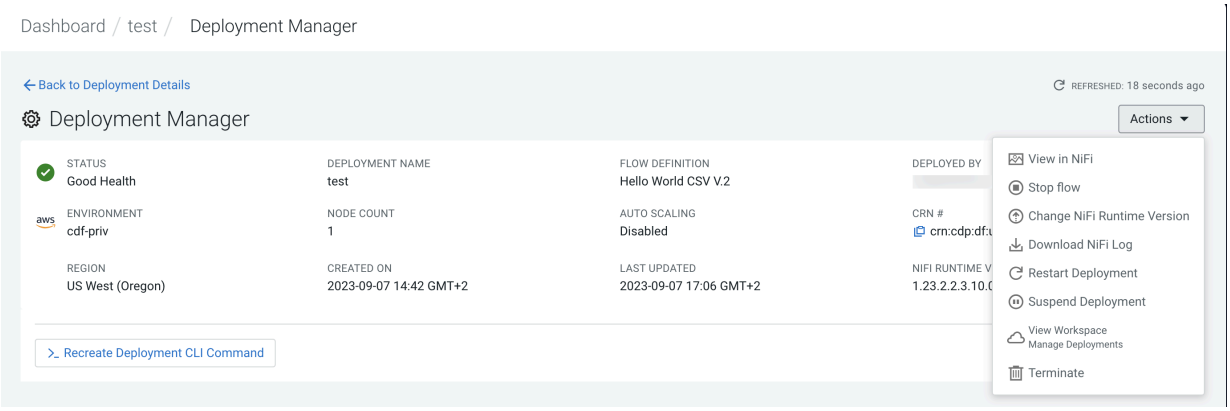
- The NiFi flow stops processing data and all underlying cloud resources are terminated. Any unprocessed data in the flow is stored in memory and its processing resumes when you resume the deployment.
- Flow persistence is maintained while a deployment is suspended.
- You cannot modify deployment configuration 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. Click Actions in the **Deployment Manager** page.



2. Click Suspend Deployment.

   The Suspend [Deployment Name] page appears.
3. Click Suspend Deployment to suspend the DataFlow deployment.

Resuming a deployment

You can resume a suspended DataFlow deployment.

About this task

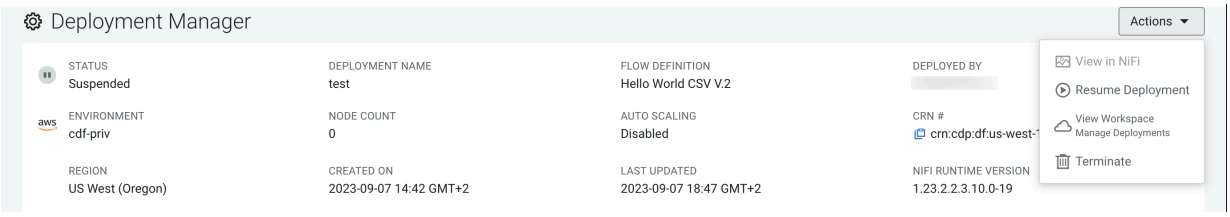
Resuming a DataFlow deployment reallocates the underlying cloud resources and returns a deployment to the state it was in before being suspended.

Before you begin

You must have a suspended flow deployment in Cloudera DataFlow.

Procedure

1. Click Actions.



2. Click Resume Deployment.

   The Resume [Deployment Name] pop-up appears.

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

Export deployment configuration

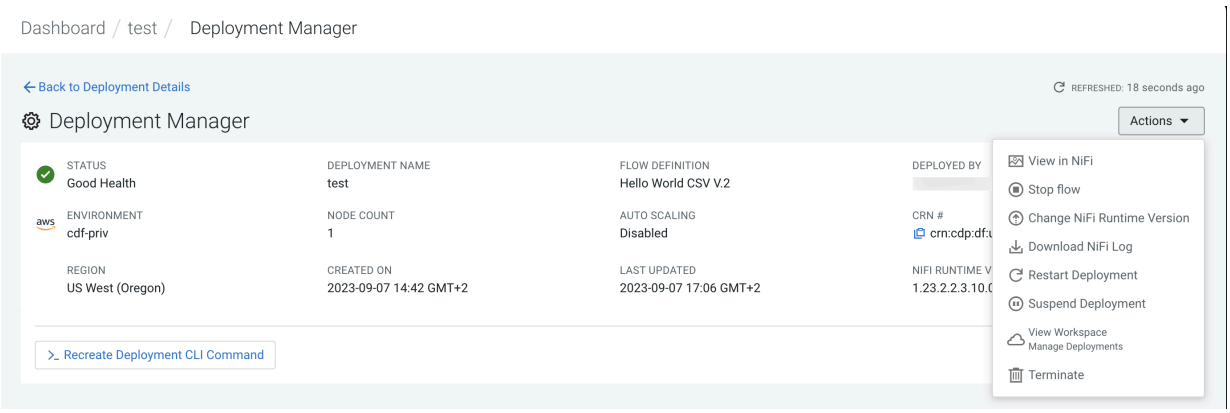
You can export a deployment configuration to create additional deployments with a similar configuration in the same or a different environment.

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. Click Actions in the **Deployment Manager** page.



2. Click Export Configuration.
3. Confirm your choice by clicking Export in the pop-up.

Results

The configuration is exported to the {LOG location}/cdf-deployment-backup directory. {LOG location} is configured during the creation of the associated CDP 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 from Cloudera DataFlow.

About this task

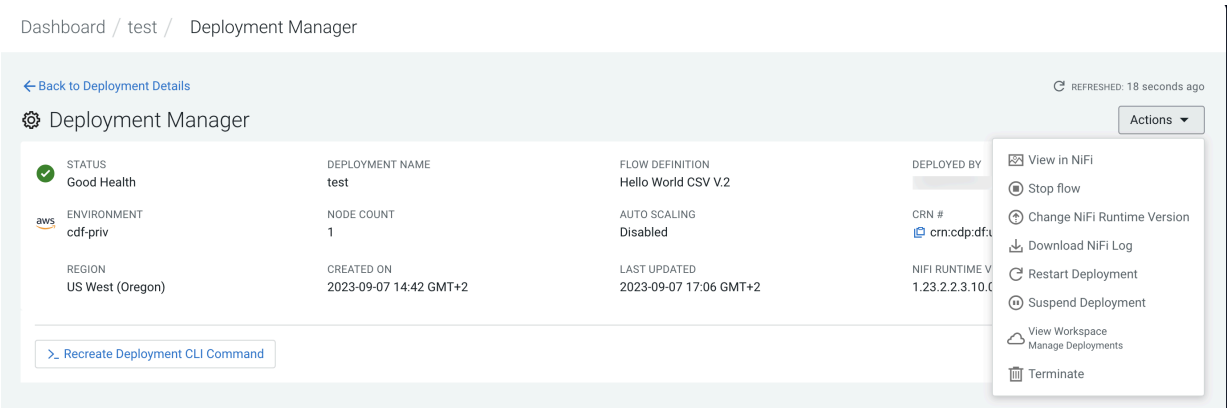
If you terminate a deployment, you delete the associated NiFi resources and your flow no longer remains active. The associated flow definition remains in the catalog and is available to be deployed again in a new deployment.

Before you begin

You must have deployed a flow definition in Cloudera DataFlow.

Procedure

1. Click Actions in the **Deployment Manager** page.



2. Click Terminate.

   The Terminate [Deployment Name] page appears.

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

   Alternately, click the clipboard icon to copy the deployment name and paste the name of the deployment, and then click Terminate.

Editing a deployment

You can edit the KPIs, alerts, parameters, size and scale details of your flow deployment under Deployment Settings or using the CDP CLI.

Before you begin

- Required Role: DFFlowAdmin
- You must have deployed a flow definition in Cloudera DataFlow.

For UI

Steps

1. Go to Deployment Manager Deployment Settings .
2. Select KPIs and Alerts to edit KPIs or the alerts they trigger.
3. Select Sizing and Scaling
   - to change the size and number of NiFi nodes provisioned for your flow deployment
   - to toggle auto-scaling
   - to view storage type
4. Select Parameters to change flow deployment parameter values.
5. Click Apply Changes before exiting each tab.

Tip:

Check the UI for CLI usage examples to help you learn to automate your common DataFlow tasks.

Result

You can confirm by finding the Deployment Update success, in Deployment Alerts.

For CLI

Before you begin

- You have installed CDP CLI.
- You have run df list-deployments. The output includes the crn field containing the deployment-crn and the service.crn field containing the environment-crn.
- To obtain the configuration-version, you have run the following command, the output of which contains the configurationVersion field containing the configuration-version value:

```
dfworkload get-deployment-configuration
--environment-crn [***ENVIRONMENT_CRN***]
--deployment-crn [***DEPLOYMENT_CRN***]
```

Steps

1. To edit an existing flow deployment, enter:

```
cdp dfworkload update-deployment
          --environment-crn <value>
          --configuration-version <value>
          --deployment-crn <value>
          [--parameter-groups <value>]
          [--auto-scaling-enabled | --no-auto-scaling-enabled]
          [--auto-scale-min-nodes <value>]
          [--auto-scale-max-nodes <value>]
          [--static-node-count <value>]
          [--kpis <value>]
          [--asset-update-request-crn <value>]
```

   - [--static-node-count] – Specifies the number of NiFi nodes when autoscaling is not enabled. You can select between 1 and 32 nodes. The default value is 1.
   - [--auto-scaling-enabled | --no-auto-scaling-enabled] – Specifies whether you want to enable autoscaling. The default is to disable autoscaling.
     - [--auto-scale-min-nodes] – Specifies the minimum nodes when you have autoscaling enabled. If you have autoscaling enabled, this parameter is required.
     - [--auto-scale-max-nodes] – Specifies the maximum nodes when autoscaling is enabled. If you have autoscaling enabled, this parameter is required.
   - [--parameter-groups] – Specifies the location of the parameter group JSON file you, if you are using one for this flow deployment.
   - [--kpis] – Specifies the location of the KPIs JSON file, if you are providing KPIs for this flow.

Example of updating your static node count

```
cdp dfworkload update-deployment \
    --environment-crn $ENVIRONMENT_CRN \
    --deployment-crn $DEPLOYMENT_CRN \
    --configuration-version $CONFIGURATION_VERSION \
    --static-node-count 2
```

Example of how to use a .json file to update parameter values

```
cdp dfworkload update-deployment \
    --environment-crn $ENVIRONMENT_CRN \
    --deployment-crn $DEPLOYMENT_CRN \
    --configuration-version $CONFIGURATION_VERSION \
    --parameter-groups file:///tmp/parameter-groups.json
```