Cloudera DataFlow

Enabling DataFlow for an Environment

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

CLOUDERA

Legal Notice

Contents

Enabling Cloudera DataFlow for an environment

Enabling Cloudera DataFlow for an environment is your first step in getting started with Cloudera DataFlow. To do this, ensure that you have met the prerequisites, and then launch the Enable Environment window to walk you through the process.

Before you begin

- You have an AWS or an Azure account.
- You have prepared your infrastructure and network. For more information on requirements, see the *AWS Resource Planning* or the *Azure Resource Planning* documentation respectively.
- You have created and registered a Cloudera Public Cloud environment. For more information, see the *Data Hub documentation*.
- FreeIPA is running and healthy.
- Your Data Lake is started and healthy.
- You have the DFAdmin role for the environment you want to enable. For more information, see the *Cloudera DataFlow Security* documentation

For UI

Steps

1. From the Cloudera Public Cloud home page, click Cloudera DataFlow, then click Environments.
2. Find the environment you want to enable, and click Enable to launch the Environments / Enable DataFlow window.

   If the Enable button is grayed out, hover over the Not Enabled icon for more details about the problem.

   Note:

   If you have previously disabled the environment but preserved event history, go to Actions and select Enable Environment.
3. Configure **DataFlow Capacity**.

   This defines the minimum and maximum size of the Kubernetes (K8s) cluster. Your Cloudera DataFlow cluster automatically scales between the minimum and maximum cluster size that you specify here.

   Note: You can specify a maximum capacity of 50 K8s nodes.

4. Configure **Networking**.

   a. Specify whether to use a Public Endpoint.

   Select this option when you want to allow users to connect to workload side UIs like the Cloudera DataFlow Deployment Manager or the actual NiFi UI through the public Internet.

   - If checked when enabling an AWS environment, this option provisions an endpoint (load balancer) in a public subnet.

     Note: A public endpoint requires public subnets. If you want to place a public endpoint in private subnets, you must explicitly select them.

     If checked when enabling an Azure environment, this option provisions an endpoint (load balancer) in a public subnet and you cannot explicitly specify subnets. The Load Balancer Subnet Use option is not available.

   - If unchecked, Cloudera Public Cloud creates an endpoint in a private subnet, and you must set up access to the endpoint manually in your cloud account to allow user access to workload side UIs.

   b. Specify Load Balancer Subnets.

   Select from Available Subnets. Explicitly specifying subnets overrides the automatic, tag-based subnet selection process and ensures load balancer provisioning in the specified subnets. If no subnets are specified, Cloudera DataFlow provisions a load balancer according to how the subnets have been tagged. For more information, see *VPC and subnets* in the *Related information* section below.

   c. Specify Worker Node Subnets.

   Select one of the Available Subnets. If you do not make a subnet selection, Cloudera DataFlow considers any available subnet that has been registered with the environment for worker placement. Worker nodes are only placed in public subnets if no private subnets are available.

   d. Specify Load Balancer Endpoint Access.

   Specify a set of IP address ranges that will be allowed to access the Cloudera DataFlow load balancer endpoint. Providing no IP address ranges makes the load balancer endpoint open to all traffic.

5. Configure **Kubernetes**.

   a. Configure API Server Endpoint Access.

   Cloudera Public Cloud environments with Cluster Connectivity Manager enabled support the ability to create a fully private cluster, which disallows all access to the Kubernetes API Server Endpoint.

   - Specify whether to use a Private Cluster.

     If you select Private Cluster IP based access is not applicable, and the option to specify a set of IP address ranges is not available.

     If you do not select Private Cluster you can specify a set of IP address ranges that are allowed to access the Kubernetes API Server Endpoint. Providing no Classless Inter-Domain Routing (CIDR) makes the Kubernetes API Server Endpoint open to all traffic.

     In either case, any user who needs access to the Kubernetes API Server must be granted remote access to the underlying Kubernetes cluster. This can be configured after Cloudera DataFlow has been enabled successfully.

     Note: If you select Private Cluster while enabling an Azure environment, the UDR Mode option becomes available. If you enable UDR mode, make sure that under Networking Worker Node Subnets you selected the specific worker node subnet where UDR mode has been configured.

   b. Configure Pod CIDR Range.

   The CIDR notation IP range from which to assign IPs to pods in your kubernetes cluster. This address should be a large address space that is not in use elsewhere in your network environment, accommodating the number of nodes that you expect to scale up to. The default value is 10.244.0.0/16. The value is used to assign a /24 address space to each node in the cluster. For example, the first node would be assigned 10.2

44.0.0/24, the second node 10.244.1.0/24, the third node 10.244.2.0/24. As the cluster scales or upgrades, the platform continues to assign a pod IP address range to each new node.

    c. Configure Service CIDR Range.

    The CIDR notation IP range from which to assign IPs to internal services in your kubernetes cluster. This IP address range should be an address space that is not in use elsewhere in your network environment, accommodating the amount of kubernetes services that you intend to use. The default value is 10.0.0.0/16.

6. Configure Tags as keys and their values.

Tags are added to Cloudera DataFlow resources at the time of enablement. These tags are included in addition to those set by the Cloudera Public Cloud service.

7. Configure Notifications.

Select the types of events for which you want to receive notifications on the Cloudera Management Console.

DataFlow Service

    If you enable this option, you will receive notifications about enablement progress, scaling operations, and health for this Cloudera DataFlow service. Select the event severity levels you want to receive notifications for (Info, Warning, Error). Cloudera recommends you keep notifications enabled for all event severities.

All Deployments

    If you enable this option, you will receive notifications for deployment progress, scaling operations, and health for all deployments in this workspace. Select the event severity levels you want to receive notifications for (Info, Warning, Error). Cloudera recommends you keep notifications enabled for all event severities.

You can configure email notifications for the events you have subscribed to after the environment is enabled for Cloudera DataFlow.

8. Click Enable. This may take up to 45 minutes.

Result

Your cluster status changes from Not Enabled to Enabling.

- Hover over Enabling for environment enablement event messages to display.
- Click the Alerts tab to see environment enablement event messages.
- Click anywhere in your environment row to see your environment details.

For CLI

Before you begin

- You have installed the Cloudera CLI.
- You have run cdp environments list-environments # to obtain the environment-crn.

Steps

1. To enable Cloudera DataFlow for an environment, enter:

```
cdp df enable-service
--environment-crn [***ENVIRONMENT_CRN***]
--min-k8s-node-count [***MIN_K8S_NODE_COUNT***]
--max-k8s-node-count [***MAX_K8S_NODE_COUNT***]
[--use-public-load-balancer] [--no-use-public-load-balancer]
[--private-cluster] [--no-private-cluster]
[--kube-api-authorized-ip-ranges
 [[***KUBE_API_AUTHORIZED_IP_RANGES***[ [[***KUBE_API_AUTHORIZED_IP_RANGES***[ ..
.]]]
[--tags [***TAGS***]]
[--load-balancer-authorized-ip-ranges
 [[***LOAD_BALANCER_AUTHORIZED_IP_RANGES***]
 [[***LOAD_BALANCER_AUTHORIZED_IP_RANGES***] ...]]]
```

```
[--cluster-subnets [[***CLUSTER_SUBNETS***] [[***CLUSTER_SUBNETS***] ...
]]]
[--load-balancer-subnets [[***LOAD_BALANCER_SUBNETS***]
 [[***LOAD_BALANCER_SUBNETS***] ...]]]
[--send-events-to-notification] [--no-send-events-to-notification]
[ --notification-severity-levels [[***NOTIFICATION_SEVERITY_LEVELS***]
 [[***NOTIFICATION_SEVERITY_LEVELS***] ...]]]
[--send-collective-flows-events-to-notification][--no-send-collective-fl
ows-events-to-notification]
 [--collective-flows-notification-severity-levels
 [[***COLLECTIVE_FLOWS_NOTIFICATION_SEVERITY_LEVELS***]
 [[***COLLECTIVE_FLOWS_NOTIFICATION_SEVERITY_LEVELS***] ...]]]
```

Where:

- --environment-crn specifies the environment-crn you obtained while completing the pre-requisites.
- --min-k8s-node-count and --max-k8s-node-count specify your Cloudera DataFlow capacity. Your Cloudera DataFlow cluster automatically scales between the minimum and maximum cluster size that you specify here.
- [--use-public-load-balancer] [--no-use-public-load-balancer]
- [--private-cluster] [--no-private-cluster]
- --kube-api-authorized-ip-ranges specifies a set of IP address ranges that will be allowed to access the Kubernetes API Server Endpoint. Providing no IP address ranges makes the Kubernetes API Server Endpoint open to all traffic.
- --tags specifies any tags you want to add to Cloudera DataFlow resources at the time of enablement. These tags are in addition to those set by the Cloudera service.
- --load-balancer-authorized-ip-ranges
- --cluster-subnets
- --load-balancer-subnets
- [--send-events-to-notification], [--no-send-events-to-notification] define whether you want to receive notifications on Cloudera Management Console about environment level events and --notification-severity-levels defines the types of events you want to receive.
- [--send-collective-flows-events-to-notification], [--no-send-collective-flows-events-to-notification] define whether you want to receive event notifications about all deployments in the environment and --collective-flows-notification-severity-levels defines the types of events you want to receive.

Example

Successfully enabling Cloudera DataFlow for an AWS environment results in output similar to:

```
{
  "service": {
    "kubeApiAuthorizedIpRanges": [],
    "validActions": [
      "ENABLE"
    ],
    "loadBalancerAuthorizedIpRanges": [],
    "clusterSubnets": [],
    "loadBalancerSubnets": [],
    "tags": {},
    "crn": "crn:cdp:df:us-west-1:CLOUDERA:service:96827a6b-bd8b-42fb-85ad-
bdb4fc7ca43d",
    "environmentCrn": "crn:cdp:environments:us-west-1:CLOUDERA:environment
:dev-east-local-k8s",
    "name": "dev-east-local-k8s",
    "cloudPlatform": "LOCAL_K8S",
    "region": "local-k8s",
    "deploymentCount": 0,
    "minK8sNodeCount": 3,
    "maxK8sNodeCount": 4,
```

```
      "status": {
        "state": "ENABLING",
        "message": "Enabling DataFlow crn:cdp:environments:us-west-1:CLOUDER
A:environment:dev-east-local-k8s",
        "detailedState": "NEW"
      },
      "runningK8sNodeCount": 0,
      "instanceType": "m5.xlarge",
      "dfLocalUrl": "",
      "activeWarningAlertCount": "0",
      "activeErrorAlertCount": "0",
      "clusterId": "",
      "clusterUsable": false,
      "usePublicLoadBalancer": false,
      "usePrivateCluster": false,
      "creatingK8sNodeCount": 0,
      "terminatingK8sNodeCount": 0
  }
}
```

What to do next

After the environment is enabled for Cloudera DataFlow, you can subscribe to receiving email notifications on events you have configured during enablement. For more information, see the Cloudera Management Console documentation.

Once you have enabled your Cloudera DataFlow environment, you are ready to deploy your first flow definition from the Catalog. For instructions, see *Deploying a flow definition*.

For more information on managing and monitoring Cloudera DataFlow, see *Managing Cloudera DataFlow in an environment*.

Related Information

AWS Resource Planning

Azure Resource Planning

Data Hub documentation

Cloudera DataFlow security

VPC and subnets in AWS environments

VNet and subnets in Azure environments

Deploying a flow definition

Managing Cloudera DataFlow in an environment

Setting up service and deployment notifications

You can configure to send notifications about Cloudera DataFlow service and deployment events to the Cloudera Management Console. You can also subscribe to email notifications on these events.

You can enable notifications for a Cloudera DataFlow service, and for all deployments within that service. These notifications are independent of each other, you can configure just one or both of them. You can either configure notifications while enabling Cloudera DataFlow for an environment, or at a later time, for a running environment.

Configuring notifications only enables Cloudera DataFlow to send the configured event type notifications to the Cloudera Management Console that you can view in-app. If you want to receive email announcements of these notifications, you need to additionally set up a subscription.

Both the description of the notification in the Cloudera Management Console and the email announcement contains a URL that allows you to easily navigate to your Cloudera DataFlow deployment or service.

Note:

The URL only works for you if you have access to the resource that the alert originated from.

Related Information

Receiving notifications | Management Console

Setting up notifications while enabling Cloudera DataFlow for an environment

Learn how you can configure to send notifications of Cloudera DataFlow service and deployment events to the Cloudera Management Console while enabling Cloudera DataFlow for an environment.

Before you begin

To set up notifications you must have the DFAdmin role for the environment.

Procedure

- During the environment enablement process, select the types of events for which you want to receive notifications on the Cloudera Management Console.

DataFlow Service

> If you enable this option, you will receive notifications about all other alerts related to this Cloudera DataFlow service. Select the event severity levels you want to receive notifications for (Info, Warning, Error).

All Deployments

> If you enable this option, you will receive notifications for all the alerts related to all of the NiFi flow deployments in this Cloudera DataFlow service. Select the event severity levels you want to receive notifications for (Info, Warning, Error).

What to do next

After the environment is enabled for Cloudera DataFlow, you can configure receiving email notifications.

Setting up notifications in a running service

Learn how you can configure to send notifications of Cloudera DataFlow service and deployment events to the Cloudera Management Console in a running Cloudera DataFlow service.

Before you begin

You have the DFAdmin role for the environment.

Procedure

1. In Cloudera DataFlow, from the **Environments** page, select the enabled Cloudera DataFlow service where you want to set up notifications.
2. Click Manage DataFlow from the **Environment Details** pane.

    You are redirected to the **Manage DataFlow** page.
3. Under **DataFlow Settings**, select the **Notifications** tab.

4. Select the types of events for which you want to receive notifications on the Cloudera Management Console.

   DataFlow Service

   > If you enable this option, you will receive notifications about all other alerts related to this Cloudera DataFlow environment. Select the event severity levels you want to receive notifications for (Info, Warning, Error).

   All Deployments

   > If you enable this option, you will receive notifications for all the alerts related to all of the Apache NiFi flow deployments in this environment. Select the event severity levels you want to receive notifications for (Info, Warning, Error).

5. Click ⬆ Apply Changes.

6. Click 🔔 Create Subscription ↗ if you want to create an email subscription now.

   You are redirected to the Cloudera Management Console where you can set up and configure your email subscription.

Setting up announcements for notifications

Learn about how to receive email announcements for events that you have subscribed to.

Before you begin

- You have the DFAdmin role for the environment.
- You have the NotificationSubscriber role for the organization.

Procedure

1. In Cloudera DataFlow, from the **Environments** page, select the environment where you want to set up notifications.
2. Click Manage DataFlow from the **Environment Details** pane.

   You are redirected to the **Manage DataFlow** page.
3. Under **DataFlow Settings**, select the **Notifications** tab.
4. Click 🔔 Create Subscription ↗.

   You are redirected to the Cloudera Management Console where you can configure email announcements.

   📝 **Note:**

   > You only receive notifications for events that occur after you have updated your preferences.

Related Tasks
Sharing a subscription link with users without the DFAdmin role

Sharing a subscription link with users without the DFAdmin role

Learn about providing access to email announcements for team members that have no access to the Environments view.

Before you begin

- To access the **Environments** view and share the subscription link, you must have the DFAdmin role for the environment.
- The users you share the subscription link with must have the NotificationSubscriber role for the organization.

Procedure

1. In Cloudera DataFlow, from the **Environments** page, select the environment for which you want to share notifications.
2. Click Manage DataFlow on the **Environment Details** pane.

   You are redirected to the **Manage DataFlow** page.
3. Under **DataFlow Settings**, select the **Notifications** tab.
4. Copy the SUBSCRIPTION LINK and share it with team members who do not have the DFAdmin role and therefore cannot access this page.

   They can subscribe to notifications as described in the Cloudera Management Console documentation.