

Troubleshooting

Date published: 2021-04-06

Date modified: 2024-01-09

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

| | |
|----------------------------------------------------------------------------------------------------|-----------|
| Overview..... | 4 |
| Collect and send diagnostic bundle using CLI..... | 4 |
| Collect and send a diagnostic bundle using Unified Diagnostics..... | 7 |
| Downloading NiFi application log..... | 10 |
| Troubleshooting errors that occur when enabling DataFlow for an environment fails..... | 11 |
| Troubleshooting errors that occur when disabling DataFlow for an environment fails..... | 15 |
| Troubleshooting errors that occur after successful DataFlow enablement..... | 15 |
| Troubleshooting Flow Deployment errors..... | 17 |

Overview

You may encounter some common errors while using Cloudera DataFlow (CDF), and it is useful to understand how to recognize and correct them.

[Troubleshooting errors that occur when enabling CDF for an environment](#)

Learn how to recognize and correct common errors that occur when you are enabling CDF for an environment.

[Troubleshooting errors that occur when CDF is disabled for an environment](#)

Learn how to recognize and correct common errors that occur when you are disabling CDF for an environment.

[Troubleshooting errors that occur when CDF is enabled for an environment](#)

Learn how to recognize and correct common errors with environments for which CDF has been enabled.

[Troubleshooting flow deployment errors](#)

Learn how to recognize and address common errors with your CDF flow deployments.

Collect and send diagnostic bundle using CLI

Learn about collecting diagnostic logs and uploading them using the CLI to facilitate troubleshooting both environment and deployment issues.

```
cdp df start-get-diagnostics-collection \
--df-service-crn '[***DATAFLOW SERVICE CRN***]' \
--description 'some description' \
--destination 'CLOUD_STORAGE' \
--environment-components '["CFM_OPERATOR", "CERT_MANAGER"]' \
--start-time [***YYYY-MM-DDTHH:MM***] \
--end-time [***YYYY-MM-DDTHH:MM***] \
--deployments '["[***CRN1***]", "[***CRN2***]"]' \
--collection-scope ALL
```

--df-service-crn

Description

Provide the Cloud Resource Name (CRN) of the DataFlow service for which you want to collect diagnostic data.

Possible values

Valid CRN

Importance

Required

--description

Description

Provide a free text description of the case for which diagnostic data is being collected. The provided description is persisted in the database and returned in the listing for future reference.

Possible values

Free text

Importance

Required

--destination

Description

Specify the upload destination for the generated bundle.

Possible values

- CLOUD_STORAGE - uploads the bundle to cloud storage associated with the DataFlow Service environment. A bundle path is included in the listing output with the format <bucket/container>/<datalake-name>/cdf/....
- SUPPORT - uploads bundle to the UDX backend where the bundle is associated with the provided case number and is automatically attached to the case.

Importance

Required

--environment-components**Description**

Specify a list of DataFlow Environment Components for which you want to collect logs. This list operates as an allowed list when present, only collecting logs for specified components. If this list is not present, it implies log collection from all components.

Possible values

| Environment component | Namespace |
|-----------------------|---------------------------|
| CFM_OPERATOR | cfm-operator-system |
| CADENCE | cadence |
| CERT_MANAGER | cert-manager |
| DFX_LOCAL | dfx-local |
| DFX_LOGGING | dfx-logging |
| FLUXCD | flux-system |
| NFS_PROVISIONER | nfs-provisioner-system |
| NGINX | nginx-ingress |
| REDIS | redis |
| VAULT | vault |
| VPA | vpa |
| ZOOKEEPER_OPERATOR | zookeeper-operator-system |

Importance

Optional

--start-time**Description**

Specify the time from which you want logs to be collected. No logs generated prior to the start time are collected.

If no end or start time is specified, log collection defaults to logs from the past 24 hours. Time zone is assumed to be UTC. Collection throws an error and does not begin if the provided end time is earlier than the provided start time.

Possible values

Date and time in *YYYY-MM-DDThh:mm* format. For example, 2023-02-03T11:00

Importance

Optional

--end-time**Description**

Specify the time to which you want logs to be collected. No logs generated after the end time are collected.

If no end or start time is specified, log collection defaults to logs from the past 24 hours. Time zone is assumed to be UTC. Collection throws an error and does not begin if the provided end time is earlier than the provided start time.

Possible values

Date and time in *YYYY-MM-DDThh:mm* format. For example, 2023-02-03T11:01

Importance

Optional

--deployments**Description**

Specify a list of DataFlow Deployment CRNs for which logs should be collected. This list operates as an allowed list when present, only collecting logs for specified deployments. If this list is not present, it allows log collection from all Flow Deployments in the environment.

Possible values

List of valid CRNs

Importance

Optional

--collection-scope**Description**

Specify the scope of data collection.

Possible values

| | |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ALL | collects logs from environment components and deployments, honoring the provided allowed lists. |
| ENVIRONMENT | collects logs from environment components, honoring the environment-components allowed list, and no logs from deployments except for those specified in the deployments allowed list. |
| DEPLOYMENT | collects logs from deployments, honoring the deployments allowed list, and no logs from environment components except for those specified in the environment-components allowed list. |

Importance

Optional

Default collection:

```
cdp df start-get-diagnostics-collection \
--df-service-crn '[***DATAFLOW SERVICE CRN***]' \
--description 'some description' \
--destination 'CLOUD_STORAGE'
```

Support upload:

```
cdp df start-get-diagnostics-collection \  
--df-service-crn '[***DATAFLOW SERVICE CRN***]' \  
--description 'some description' \  
--destination 'SUPPORT' \  
--case-number '123456'
```

Note that case-number is required when destination is SUPPORT

List collection attempts:

```
cdp df list-diagnostics \  
--df-service-crn '[***DATAFLOW SERVICE CRN***]'
```

Related Tasks

[Collect and send a diagnostic bundle using Unified Diagnostics](#)

Related Information

[Send a diagnostic bundle using the CDP Management Console](#)

Collect and send a diagnostic bundle using Unified Diagnostics

You can trigger diagnostic bundle collection for your CDP environment in order to send Data Lake, FreeIPA, and DataFlow service and environment logs to Cloudera Support for support case resolution.

Before you begin

Required role: You need the EnvironmentAdmin or Owner role for the environment for which you would like to create a bundle.

- FreeIPA deployment time must be September 2020 or later.
- The VMs from which diagnostic data is to be collected must be running.
- Salt minions must be started properly on the Management Console nodes.

Procedure

1. Log in to the CDP web interface.
2. Navigate to the Management Console.

- Click Help in the bottom left corner and select Collect Diagnostic Data.

The screenshot shows the Cloudera Management Console interface. On the left is a sidebar with navigation links: Dashboard, Environments, Data Lakes, User Management, Data Hub Clusters, Data Warehouses, ML Workspaces, Classic Clusters, Data Services Clusters, Audit, Consumption, Shared Resources, Global Settings, and Notifications. Below these is a 'Get Started' section with a 'Help' link highlighted by a green arrow. The main content area is titled 'Help' and contains two sections: 'Environments' and 'Helpful Links'. The 'Environments' section explains the concept of an environment in CDP and lists related links for AWS, Azure, and GCP. The 'Helpful Links' section includes links for Support, Collect Diagnostic Data (highlighted by a green arrow), Documentation, Community, and Download CLI. On the right, the 'Environments' table is visible, showing columns for Cloud Provider, Region, Data Lake, CDP Runtime Version, and Time Created. The table lists various environments with their status (e.g., Not registered, Creating Stack, Running, Provisioning Failed, Deleted on provider).

- Click the Collect New Diagnostic Data.
- On the **General Information** page, provide the following:

Environment

Select the environment for which you would like to generate a diagnostic bundle. This is a required field.

Case Number

Enter the related support case number. You can obtain the case number from the MyCloudera portal or from your email by checking support case notification emails. This is a required field.

Description

Provide a free text description of the case for which you would like to generate diagnostic data. The provided description is persisted in the database and returned in the listing for future reference. This is a required field.

Time Range

You can specify specific points in time (date and time) from which to start and end the collection. This is an optional field.

- Click Next.

7. On the **Selected Services** page under DataFlow, you can configure the diagnostic bundles you want to generate for CDF environment components and DataFlow deployments:

Environment Components

Select a list of DataFlow Environment Components from which you want to collect logs. This list operates as an allowed list when present, only collecting logs for specified components.

If you leave this list empty, it implies you want to collect logs from all components.

Select Components

Specify a list of DataFlow Deployment CRNs for which you want to collect logs. This list operates as an allowed list when present, only collecting logs for specified deployments.

If you leave this list empty, it implies log collection from all Flow Deployments in the environment.

Collection Scope

Specify whether you want to collect environment logs, deployment logs, or both.

Regardless of the selection you make here, logs for environment components and deployments you specified in their respective allowed lists will be collected.

To generate diagnostic data for a specific DataFlow Deployment, enter the CRN of the DataFlow Deployment for which you want to collect logs in the Select Components field and select DEPLOYMENT as Collection Scope.



Note: To obtain the CRN of a DataFlow Deployment:

- Navigate to DataFlow on the CDP web interface.
- On the **Dashboard** select the deployment you want to collect diagnostic data for.
- Click Copy under CRN #.

DataFlow Service: cdf-priv

Collect Diagnostic Data For

Environment Components: Select Components

Select Components: crn:cdp:df:us-west-1:9d74eee4-1cad-45d7-b645-7ccf

Collection Scope: DEPLOYMENT

Buttons: Back, Collect

Cloudera will collect Data Lake and FreeIPA diagnostics by default to assist Cloudera Support troubleshooting the issue.

To obtain all diagnostic data for a CDF Service, set Collection Scope to ALL.

DataFlow Service: cdf-priv

Collect Diagnostic Data For

Environment Components: Select Components

Select Components: Add

Collection Scope: ALL

Buttons: Back, Collect

Cloudera will collect Data Lake and FreeIPA diagnostics by default to assist Cloudera Support troubleshooting the issue.

8. Click Collect.

New entries appear on the page, allowing you to track the status of the log collection. Since a separate bundle is created for each service, there is one entry for each of the following: Data Lake, FreeIPA, and one entry for each DataFlow service. While the collection is in progress, the collection status is In progress:

Diagnostics

Collected data might include metadata, configurations, and log data from the selected services. Sensitive information, including but not limit to passwords and personal identifiers are not collected. All data is collected in accordance with the Cloudera Privacy Policy.

Environment aws cdf-priv

Q Search Diagnostic Bundles Last updated a few seconds ago Collect New Diagnostic Data

| Status | Case Number | Service Type | Service Name | Collection Start |
|-------------|-------------|--------------|--------------|-------------------------|
| In Progress | 789456 | DataFlow | | 09/28/2023 9:08 PM CEST |
| In Progress | 789456 | Data Lake | | 09/28/2023 9:08 PM CEST |
| In Progress | 789456 | Free IPA | | 09/28/2023 9:08 PM CEST |

Results

Once the collection is complete, the status changes to Completed. This means that Cloudera Support can now access the diagnostic bundles:

Diagnostics

Collected data might include metadata, configurations, and log data from the selected services. Sensitive information, including but not limit to passwords and personal identifiers are not collected. All data is collected in accordance with the Cloudera Privacy Policy.

Environment aws cdf-priv

Q Search Diagnostic Bundles Last updated a few seconds ago Collect New Diagnostic Data

| Status | Case Number | Service Type | Service Name | Collection Start |
|-----------|-------------|--------------|--------------|-------------------------|
| Completed | 789456 | DataFlow | | 09/28/2023 9:08 PM CEST |
| Completed | 789456 | Data Lake | | 09/28/2023 9:08 PM CEST |
| Completed | 789456 | Free IPA | | 09/28/2023 9:08 PM CEST |

Related reference

[Collect and send diagnostic bundle using CLI](#)

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.

Troubleshooting errors that occur when enabling DataFlow for an environment fails

Learn how to recognize and correct common errors that occur when you are enabling Cloudera DataFlow (CDF) for an environment.

Your ability to deploy flow definitions depends on DataFlow being enabled and in good health in your target environment. If a DataFlow deployment is unhealthy in an environment, it is typically because enabling or disabling DataFlow for an environment has failed. Review the environment troubleshooting information to understand common errors and their solutions.

When you enable DataFlow for an existing CDP environment, CDP creates the required infrastructure in your cloud environment and installs core CDP services as well as the DataFlow software in the Kubernetes cluster. The enablement process can be divided into three parts:

1. Provisioning Cloud Infrastructure
2. Installing core CDP software and services
3. Installing CDF software and services

To understand why the enablement process is failing, the first step is to identify at which stage the enablement process failed.

Identifying where the enablement process failed

Use the information provided in the hover state of an environment and the status messages that have been logged in the environment's Event History to identify where the enablement process has failed.

1. In DataFlow, navigate to the Environments page and find the environment where the enablement process failed.
2. Hover over the status icon and take note of the error message.
3. Click the environment, select the Alerts tab, and review the error and info events that have been logged during enablement.

Enablement fails during infrastructure provisioning or core CDP software installation

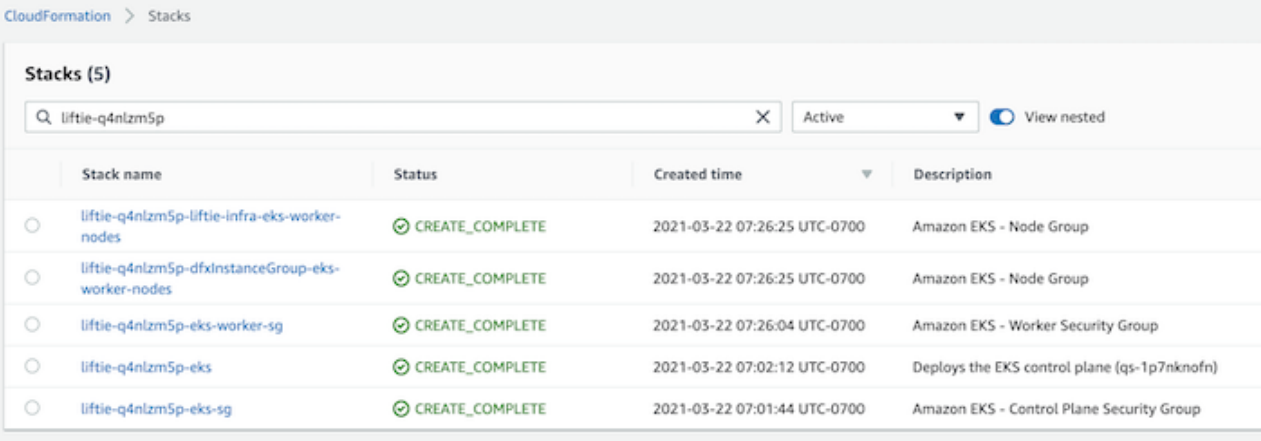
If you only see a status message in the Event History indicating that the infrastructure provisioning has started but you cannot see a corresponding status message confirming that the infrastructure has been provisioned successfully, CDP was either not able to create the required infrastructure or install the core CDP software and services afterwards.

| | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|---|
| ❗ Enable Failed | 2020-10-12 04:43 PDT | ⬆ |
| ALERT DETAILS: Failed to enable environment. Reason : [\n not found: limit=0 content=...] | | |
| DURATION: 1 seconds | | |
| ❗ Infrastructure Provisioning Failed | 2020-10-12 04:43 PDT | ⬆ |
| ALERT DETAILS: Failed to create Kubernetes cluster with ID UNKNOWN. Reason : [unexpected end of stream on http://computex-app-cpx-liftie.computex.svc.cluster.local:9999/...] | | |
| DURATION: 1 seconds | | |
| ℹ Provisioning Infrastructure | 2020-10-12 04:43 PDT | ⬇ |
| ℹ Signing Certs with UMS | 2020-10-12 04:43 PDT | ⬇ |

The infrastructure provisioning failed error message indicates that there was either an issue with creating AWS infrastructure or setting up core CDP services.

Validating infrastructure creation – for AWS users

CDP uses CloudFormation scripts to create the required infrastructure in your AWS account. To validate whether the requested resources have been created successfully, log in to your AWS account, navigate to CloudFormation and search for the Kubernetes cluster ID that you have extracted from the environment events and looks similar to `liftie-q4nlzm5p`. Verify that the CloudFormation scripts completed successfully.



| Stack name | Status | Created time | Description |
|---------------------------------------------------|-----------------|------------------------------|----------------------------------------------|
| liftie-q4nlzm5p-liftie-infra-eks-worker-nodes | CREATE_COMPLETE | 2021-03-22 07:26:25 UTC-0700 | Amazon EKS - Node Group |
| liftie-q4nlzm5p-dfxInstanceGroup-eks-worker-nodes | CREATE_COMPLETE | 2021-03-22 07:26:25 UTC-0700 | Amazon EKS - Node Group |
| liftie-q4nlzm5p-eks-worker-sg | CREATE_COMPLETE | 2021-03-22 07:26:04 UTC-0700 | Amazon EKS - Worker Security Group |
| liftie-q4nlzm5p-eks | CREATE_COMPLETE | 2021-03-22 07:02:12 UTC-0700 | Deploys the EKS control plane (qs-1p7nknofn) |
| liftie-q4nlzm5p-eks-sg | CREATE_COMPLETE | 2021-03-22 07:01:44 UTC-0700 | Amazon EKS - Control Plane Security Group |

If the CloudFormation script did not complete successfully, make sure that the cross account role for your CDP environment has been assigned appropriate permissions.

If the CloudFormation script completed successfully but enabling Dataflow failed before completing the infrastructure setup, this might be an indication that the Kubernetes cluster cannot communicate with the CDP control plane or other public endpoints like container image repositories. Make sure that the VPC and subnets you are using for DataFlow meet the CDP and DataFlow prerequisites.

Validating infrastructure creation – for Azure users

CDP uses Azure Resource Manager to orchestrate infrastructure creation in Azure. When you enable DataFlow for a CDP environment, a Cluster ID is generated that can be used to track all associated resources in Azure. Navigate to Environments in CDF, select your Azure environment and copy the Cluster ID, which will look similar to `liftie-fj6hq9sc`.

»

✓ dataflow-azure
West US 2

[Details](#) Alerts

DataFlow Information

KUBERNETES NODE ALLOCATION
15% (3 of 20)

LAST UPDATED
2022-01-24 14:23 PST

DATAFLOW VERSION
2.0.0-b6

CLUSTER ID
liftie-fj6hq9sc

DATAFLOW CRN

crn:cdp:df:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb73d:service:15af86e9-e707-498a-9a7d-c4112111f2af

First, make sure that CDF's Azure PostgreSQL database has been created successfully. In the Azure Portal, navigate to your CDP resource group, explore Deployments under Settings, and find the database deployment associated with the previously obtained Cluster ID.

Microsoft Azure

Search resources, services, and docs (G+)

Home > mikahs-rg

-rg | Deployments

Search (Cmd+J)

Refresh Cancel Redeploy Delete View template

Filter by deployment name or resources in the deployment...

Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Events

Settings

Deployments

Security

Policies

Properties

Locks

Cost Management

| Deployment name | Status | Last modified | Duration | Related events |
|----------------------------------|-----------|-------------------------|----------------------|----------------|
| df-db-stack-liftie-fj6hq9sc30599 | Succeeded | 1/24/2022, 12:51:16 AM | 2 minutes 48 seconds | Related events |
| df-db-stack-liftie-hunktfu30464 | Succeeded | 1/20/2022, 4:52:14 AM | 3 minutes 43 seconds | Related events |
| df-db-stack-liftie-5tpffbs630363 | Succeeded | 1/18/2022, 11:45:56 AM | 2 minutes 47 seconds | Related events |
| df-db-stack-liftie-ctf6j9t830271 | Succeeded | 1/17/2022, 12:20:36 AM | 3 minutes 1 second | Related events |
| df-db-stack-liftie-qg85hbd30192 | Succeeded | 1/13/2022, 7:20:07 AM | 2 minutes 47 seconds | Related events |
| df-db-stack-liftie-cbqdmmlw30085 | Succeeded | 1/11/2022, 11:53:09 AM | 2 minutes 51 seconds | Related events |
| df-db-stack-liftie-kcznd85k28916 | Succeeded | 12/6/2021, 1:53:06 PM | 3 minutes 48 seconds | Related events |
| df-db-stack-liftie-hdttfcr28705 | Succeeded | 12/1/2021, 12:03:54 PM | 2 minutes 46 seconds | Related events |
| df-db-stack-liftie-k5sqjps28454 | Succeeded | 11/23/2021, 10:09:48 AM | 3 minutes 44 seconds | Related events |
| df-db-stack-liftie-ts82hj228412 | Succeeded | 11/22/2021, 9:58:39 AM | 2 minutes 47 seconds | Related events |
| df-db-stack-liftie-kwcf62a28411 | Succeeded | 11/22/2021, 9:40:35 AM | 2 minutes 46 seconds | Related events |

Next, validate that the AKS Kubernetes cluster and associated infrastructure has been created successfully. In the Azure Portal, use the Cluster ID to search for associated resources in your Azure subscription.

liftie-fj6hq9sc

Services

No results were found.

Marketplace

No results were found.

Resources

liftie-fj6hq9sc

Kubernetes service

liftie-fj6hq9sc-a8a30f9b-5321-4fd2-9c15-74...

Log Analytics workspace

Documentation

No results were found.

Resource Groups

MC_liftie-fj6hq9sc_westus2












You should find a Kubernetes service as well as a new resource group called MC_<Cluster ID>_<AZURE Region>, which contains the virtual machine scale sets, load balancers, IP addresses, and storage disks that have been created for the Kubernetes cluster.

Resources Recommendations

Type == **all** ✕
Location == **all** ✕
+ Add filter

Showing 1 to 11 of 11 records. ☐ Show hidden types ⓘ

☐ **Name** ↑↓

| | |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> |  875728c7-eaff-4d68-ae3b-c9a173f36041 |
| <input type="checkbox"/> |  aks-agentpool-82520391-nsg |
| <input type="checkbox"/> |  aks-agentpool-82520391-routetable |
| <input type="checkbox"/> |  aks-dfinfra-29645858-vmss |
| <input type="checkbox"/> |  aks-liftieinfra-29645858-vmss |
| <input type="checkbox"/> |  kubernetes |
| <input type="checkbox"/> |  kubernetes-internal |
| <input type="checkbox"/> |  pvc-3ac204b0-09e6-432b-8805-9fadac65b712 |
| <input type="checkbox"/> |  pvc-5821631e-ec7e-4301-8a90-87c1f8d6306d |
| <input type="checkbox"/> |  pvc-8f1b3a6e-4d17-4f78-80f2-0bcf3b2eca9d |
| <input type="checkbox"/> |  pvc-be306101-4849-4a4a-9f97-71eccc62ee58 |

If the resources were not created successfully, make sure that the app-based credential for your CDP environment has been assigned the appropriate permissions. For more information, see *Prerequisites for the provisioning credential*.

If the infrastructure resources were created successfully, but enabling DataFlow failed before completing the infrastructure setup, it might be an indication that the Kubernetes cluster cannot communicate with the CDP control plane or other public endpoints like container image repositories. Make sure that the virtual network and subnets you are using for DataFlow meet the CDP and DataFlow prerequisites. For more information, see *VNet and subnets*.

Enablement fails during CDF software and service installation

If you see a status message that indicates that the required Infrastructure has been provisioned successfully but the enablement process still failed, this is an indication that installing and setting up the CDF software and services has failed.

| | | |
|-------------------------------|----------------------|---|
| ① Infrastructure Provisioned | 2021-03-22 07:36 PDT | ▼ |
| ① Cert Signing Complete | 2021-03-22 07:02 PDT | ▼ |
| ① Provisioning Infrastructure | 2021-03-22 07:01 PDT | ▼ |
| ① Signing Certs with UMS | 2021-03-22 07:01 PDT | ▼ |
| ① Enable Initiated | 2021-03-22 07:01 PDT | ▼ |

The Infrastructure Provisioned status event indicates that the Kubernetes cluster has been created and core CDP services have been setup successfully.

To ensure that this is not a transient issue, use the Retry Enablement action to start the enablement process again. Retry Enablement terminates all existing resources and provisions new infrastructure.

If retrying does not help and enabling DataFlow still fails after successfully provisioning the infrastructure, copy the error message from the Event History and open a support case with Cloudera.

Related Concepts

[Overview](#)

Related Information

[Prerequisites for the provisioning credential](#)

[VNet and subnets](#)

Troubleshooting errors that occur when disabling DataFlow for an environment fails

Learn how to recognize and correct common errors that occur when you are disabling Cloudera DataFlow (CDF) for an environment.

DataFlow can be in Bad Health due to a failed disablement process. If you see error messages in the Event History indicating that disabling DataFlow has failed, use the Retry Disable Process to try again to ensure that the issue is not transient.

| | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|---|
| ① Disable Failed | 2020-10-13 23:08 PDT | ▲ |
| ALERT DETAILS: Failed to disable Environment. Reason : [Activity 'undeployNifiDependencies' will be retried due to ['Failed to undeploy NiFi cluster dependencies from cluster with ID llftie-f9bn98km'].] DURATION: 1 seconds | | |

If the issue persists, select **Environments Action Reset Environment** to clear DataFlow's state for the environment. Resetting the Environment allows you to enable DataFlow for the same environment again at a later stage.



Note:

Resetting DataFlow does not terminate the associated cloud infrastructure. After resetting DataFlow for an environment, you must manually delete the cloud resources that were created during DataFlow enablement.

Related Concepts

[Overview](#)

Troubleshooting errors that occur after successful DataFlow enablement




Learn how to recognize and correct common errors with environments for which Cloudera DataFlow (CDF) has been enabled.

After you have successfully enabled DataFlow for an environment there are several reasons why DataFlow's health can become Concerning or Bad.

Concerning health due to "Workload Failed to Heartbeat"

If DataFlow health is concerning for one of your environments, hover over the status icon and check the Alerts tab in the environment details to see details about the issue. Workload Failed to Heartbeat means that the Cloudera Control Plane has not received a recent heartbeat from the DataFlow workload application running in your cloud account.

Active Alerts [?](#)

| Alert Type | Alert Time ↓ |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
|  Workload Failed to Heartbeat ALERT DETAILS: Failed to receive any recent heartbeats from workload. Learn more  FIRST OCCURRED: 2021-03-09 11:13 PST | 14 days ago  |

DataFlow fails to receive heartbeats from a particular environment

Heartbeat failures can have several reasons. Make sure that:

- The associated CDP environment has been started and is running.
- There was no networking related change in your VPC/subnet configuration and that your networking setup still meets the requirements outlined in the *DataFlow Networking*.

If the issue persists, open a support case with Cloudera.

Concerning Health due to Nearing Maximum Kubernetes Limit

When your Kubernetes cluster is close to its maximum node count, DataFlow will show Concerning Health for the particular environment and display an Active Alert with more details about the boundaries. To return DataFlow to Good Health you can adjust the maximum node number through the Edit Configuration option in the Environment actions menu.

Bad Health due to issues with the associated CDP environment

Certain issues with the associated CDP environment will result in DataFlow reporting Bad Health for an environment. Once DataFlow is in Bad Health you can no longer create or terminate flow deployments in the environment.

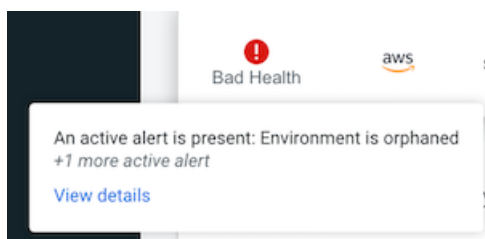
Bad Health due to CDP environment state

If the associated CDP environment is either unhealthy or in a starting/stopping state, DataFlow will report Bad Health. To return DataFlow to Good Health make sure that:

- The associated CDP environment has been started and is running.
- FreeIPA and DataLake are both running.

Bad Health due to CDP environment having been deleted

If the associated CDP environment has been deleted without disabling DataFlow first, DataFlow will report Bad Health indicating that it has been orphaned and required CDP services such as FreeIPA are no longer available in the environment.



You are not able to create new Flow Deployments for DataFlow. You cannot recover DataFlow health in this situation. Terminate your Flow Deployments and use the **Disable Environment** action to terminate DataFlow and associated cloud infrastructure and enable DataFlow again for a different CDP environment.

Related Concepts

[Overview](#)

Troubleshooting Flow Deployment errors

Learn how to recognize and address common errors with your Cloudera DataFlow (CDF) flow deployments.

Setting up `kubectl` to connect to the DataFlow Kubernetes cluster

It is helpful to have access to the DataFlow Kubernetes cluster using command line tools such as `kubectl` when you are troubleshooting deployment or upgrade failures. To set up `kubectl` access, follow these steps:

1. In DataFlow, from the Environments page, select the DataFlow service 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. Add the AWS IAM role that you will authenticate as to the list of authorized users for DataFlow by entering the ARN in the **Add User** dialog.
5. Use the **Download Kubeconfig** action to retrieve the kubeconfig file for connecting to your cluster.
6. Set up your `kubectl` to use the downloaded kubeconfig file:

```
export KUBECONFIG=[ ***PATH/TO/DOWNLOADED/KUBECONFIG/FILE*** ]
```

7. Run `kubectl get ns` and validate that your output looks similar to:

| NAME | STATUS | AGE |
|----------------------------|--------|-----|
| cadence | Active | 37h |
| cert-manager | Active | 37h |
| cfm-operator-system | Active | 37h |
| default | Active | 38h |
| dfx-dev-environment-ns | Active | 36h |
| dfx-idbrokers3-ns | Active | 36h |
| dfx-kafkatos3-ns | Active | 22h |
| dfx-local | Active | 37h |
| dfx-ops | Active | 37h |
| kube-node-lease | Active | 38h |
| kube-public | Active | 38h |
| kube-system | Active | 38h |
| liftie | Active | 37h |
| logging | Active | 37h |
| monitoring | Active | 37h |
| nfs-provisioner-system | Active | 37h |
| nginx-ingress | Active | 37h |
| prometheus-operator-system | Active | 37h |

With `kubectl` being set up correctly, you are able to access NiFi and other DataFlow logs directly through the CLI.








Understanding flow deployment failures

The Flow Deployment process consists of two phases:

1. Scheduling resources on Kubernetes and creating a new NiFi cluster:

| | | |
|-------------------------------------------------------------------------------------------------------------|----------------------|---|
|  NiFi Cluster Provisioned | 2021-03-22 22:59 PDT | ▼ |
|  Provisioning NiFi Cluster | 2021-03-22 22:58 PDT | ▼ |
|  Deployment Initiated | 2021-03-22 22:58 PDT | ▼ |

2. Importing, configuring, and starting the NiFi flow definition:

| | | |
|---------------------------------------------------------------------------------------------------------|----------------------|---|
|  Deployment Successful | 2021-03-22 22:59 PDT | ▼ |
|  Alert Rules Deployed | 2021-03-22 22:59 PDT | ▼ |
|  Deploying Alert Rules | 2021-03-22 22:59 PDT | ▼ |
|  NiFi Flow Started | 2021-03-22 22:59 PDT | ▼ |
|  Starting NiFi Flow | 2021-03-22 22:59 PDT | ▼ |
|  NiFi Flow Imported | 2021-03-22 22:59 PDT | ▼ |
|  Importing NiFi Flow | 2021-03-22 22:59 PDT | ▼ |

If your flow deployment fails, check the Event History to see where exactly the issue occurred.

Deployment fails during Phase 1

If the issue occurs during Phase 1 while scheduling resources on Kubernetes and creating the NiFi cluster, you can get more details on why the deployment failed by looking at the DataFlow application logs.

Identify the DataFlow application pod by running:

```
kubectl get pods
--namespace dfx-local
```

The result should look similar to

| NAME | READY | STATUS | RESTARTS | AGE |
|---------------------------------------|-------|-----------|----------|-----|
| create-db-zxb2q | 0/1 | Completed | 0 | 38h |
| dfx-local-deployment-7f8b466c68-xwrbf | 3/3 | Running | 0 | 38h |

Copy the dfx-local-deployment pod ID and run to view the DataFlow application logs.

```
kubectl logs \
-f dfx-local-deployment-7f8b466c68-xwrbf \
-c dfx-local \
--namespace dfx-local
```



Note:

-f tails the log file as DataFlow is writing to it.

A common reason for flow deployment issues is that the Kubernetes cluster does not have enough resources available for the deployment pods to be scheduled. Pods that are stuck in pending are an indicator for not enough resources being available.

You can explore flow deployments and their resources by running:

```
get pods --namespace dfx-deployment_name-ns
```

A healthy deployment should look similar to this with one or more NiFi pods (depending on the Sizing & Scaling settings), a ZooKeeper and a Prometheus pod.

| NAME | READY | STATUS | RESTARTS | AGE |
|-------------------------|-------|---------|----------|-----|
| dfx-nifi-0 | 5/5 | Running | 0 | 22h |
| dfx-zookeeper-0 | 1/1 | Running | 0 | 22h |
| prometheus-prometheus-0 | 4/4 | Running | 0 | 22h |

If one of the pods is stuck in Pending you can explore the pod further and identify any potential issues by looking at its events.

If in the above screenshot dfx-nifi-0 was Pending and you wanted to find out why, you would run the follow command to get detailed information about the containers in the pod.

```
kubectl describe pod dfx-nifi-0 -n dfx-deployment_name-ns
```

Find the Events section and check if there are any messages about why a container could not be scheduled.

If the flow deployment failed because of insufficient resources in the Kubernetes cluster, you can increase the Kubernetes cluster size by using the Edit Configuration action of the affected environment.

Deployment fails during Phase 2

If the issue occurs during Phase 2, check the NiFi canvas of the deployment for any error messages. To get there, open the deployment details, click Manage Deployment and in the following Deployment Manager page select the view in NiFi action.

If a processor or controller service failed to start, make sure that you have provided the correct values for the deployment parameters. You can adjust parameter values in the NiFi canvas and restart processors or controller services as needed. Once you have identified the issue, note down the correct parameter values and start a new deployment.

To view the NiFi log for a particular deployment run the following `kubectl` command.

```
kubectl logs -f dfx-nifi-0 -c app-log --namespace dfx-deployment_name-ns
```



Note:

`-f` tails the log file as NiFi is writing to it.

Related Concepts

[Overview](#)