

CDP Public Cloud

Diagnostic Bundle Collection

Date published: 2019-08-22

Date modified:

CLOUDERA

<https://docs.cloudera.com/>

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

Send diagnostic bundle to Cloudera.....4

CDP CLI commands..... 8

Diagnostic bundle content.....11

Send a diagnostic bundle to Cloudera Support

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



Note: Diagnostic bundles can only be generated for Data Lake, FreeIPA, Data Hub, Cloudera DataFlow (CDF), and Cloudera Data Warehouse (CDW). Additionally, there are separate interfaces for generating bundles for Cloudera Data Engineering (CDE) and Cloudera Machine Learning (CML). Other CDP services are not supported. A separate bundle is created for each cluster. To send a CML diagnostic bundle, follow the instructions in [Download diagnostic bundles for an ML workspace](#) instead of following the instructions below. To send a CDE diagnostic bundle, follow the instructions in [Downloading a diagnostic bundle for Cloudera Data Engineering](#).

Required role: You need the EnvironmentAdmin or Owner role for the environment for which you would like to create a bundle. If you would like to include CDW diagnostics in the bundle, you also need the DWAdmin role.

Prerequisites

- 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 CM nodes.

Steps

1. Log in to the CDP web interface.
2. Navigate to the Management Console.
3. 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 dark sidebar with navigation links: Dashboard, Environments, Data Lakes, User Management, Data Hub Clusters, Data Warehouses, ML Workspaces, Classic Clusters, Cost Management, and Global Settings. Below these are 'Get Started', 'Help' (highlighted with an orange arrow), and 'Dominika Bialek'. The main content area is titled 'Help' and contains sections for 'Environments' (with a description of the environment concept), 'Related Links' (AWS, Azure, GCP), and 'Helpful Links' (Support, Collect Diagnostic Data - highlighted with an orange arrow, Documentation, Community, and Download CLI). On the right, there is a table titled 'Environments' with columns: Region, Data Lake, CDP Runtime Version, and Time Created. The table lists several environments with their status (e.g., Not registered, Creating Stack, Provisioning Failed, Running).

Region	Data Lake	CDP Runtime Version	Time Created
US West (Oregon)	Not registered	missing data	7/12/2021, 10:43:43 AM PDT
US West (Oregon)	Creating Stack...	7.2.11	7/12/2021, 10:41:34 AM PDT
Texas (US Texas)	Provisioning Failed		7/12/2021, 10:21:20 AM PDT
US West (Oregon)	Not registered	missing data	7/12/2021, 10:08:13 AM PDT
US West (Oregon)	Running	7.2.10	7/12/2021, 9:34:41 AM PDT
US West (Oregon)	Running	7.2.11	7/12/2021, 9:26:29 AM PDT
West US 2	Running	7.2.11	7/12/2021, 9:20:37 AM PDT
US West (Oregon)	Running	7.2.11	7/12/2021, 9:18:27 AM PDT
Texas (US Texas)	Not registered	missing data	7/12/2021, 7:54:29 AM PDT
US West (Oregon)	Running	7.2.8	7/12/2021, 7:24:17 AM PDT
US West (Oregon)	Running	7.2.7	7/12/2021, 6:52:53 AM PDT
US West (Oregon)	Running	7.2.10	7/12/2021, 6:45:12 AM PDT
Texas (US Texas)	Not registered	missing data	7/12/2021, 6:29:44 AM PDT
US West (Oregon)	Running	7.2.10	7/12/2021, 5:33:24 AM PDT

4. Click the Collect New Diagnostic Data button.

5. On the General Information page, provide the following:

- Select the environment for which you would like to generate a diagnostic bundle. This is required.
- Enter the related support case number. This is required. You can obtain the case number from the MyCloudera portal or from your email by checking support case notification emails.
- Under Time Range, you can optionally specify a specific point in time (date and time) from which to start the collection. The cutoff date and time is automatically set to the time when you initiate the collection and it cannot be modified.

General Information

Environment

 **udx-test** AVAILABLE 

Case Number ⓘ

769355 

Time Range (Optional) ⓘ

Enter Start Time → Now 

Collection Method ⓘ

☒ Upload to Cloudera Support

6. Click Next.

7. On the Selected Services page:

- CDP automatically collects logs from all nodes of the Data Lake and FreeIPA clusters. By default diagnostics is collected from all nodes, but you can optionally limit it to one or more specific nodes.
- Under Data Hub clusters, you can optionally select to generate diagnostic bundles for Data Hub clusters. After adding a Data Hub, by default diagnostics is collected from all nodes, but you can optionally limit it to one or more specific nodes.
- Under Data Warehouse clusters, you can optionally select to generate diagnostic bundles for Hive or Impala virtual warehouses created via CDW.
- Under DataFlow, you can optionally select to generate diagnostic bundles 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.

Collect Diagnostic Data / eng-m1-dev-env-aws

Environmenteng-m1-dev-env-aws

Case Number1234

Time RangeAll Time

MethodUpload to Cloudera Support

Selected Services

▼ Data Hub Clusters

Q Search to add clusters

Add

▼ Data Warehouse Clusters

EnvironmentCollect Diagnostic Data For

eng-m1-dev-env-awsWarehousesType to select Warehouse(s)

▼ Data Lake Cluster

ClusterCollect Diagnostic Data For

eng-m1-dev-env-aws-dlHostsAll Hosts

▼ FreeIPA

EnvironmentCollect Diagnostic Data For

eng-m1-dev-env-awsHostsAll Hosts

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

Back

Collect

8. Click Collect.
9. 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, one entry for each Data Hub, one entry for each CDW warehouse, 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.

Filters

eng-m1-dev-env-aws

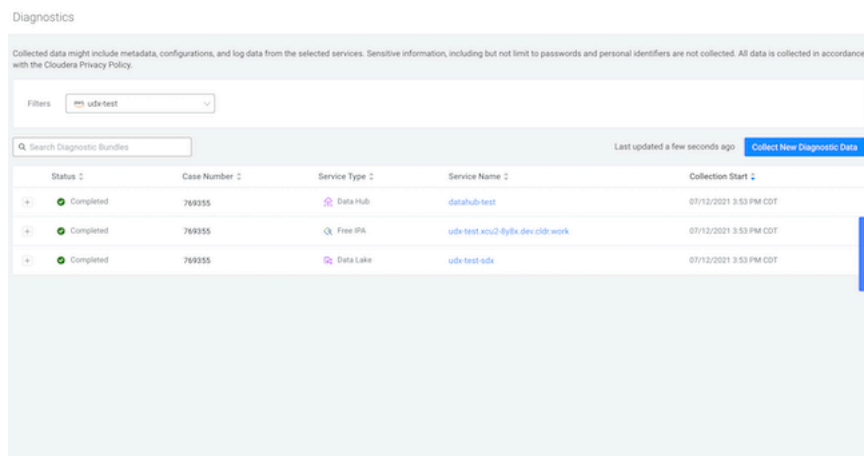
Q Search Diagnostic Bundles

Last updated a few seconds ago

Collect New Diagnostic Data

Status	Case Number	Service Type	Service Name	Collection Start
In Progress	769355	Data Hub	datahub-test	07/12/2021 3:53 PM CDT
In Progress	769355	Free IPA	udw-test.xcu2-tylxl dev.cldr-work	07/12/2021 3:53 PM CDT
In Progress	769355	Data Lake	udw-test-edx	07/12/2021 3:53 PM CDT

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



Related Information

[Diagnostic bundle content](#)

[CDP CLI commands for generating a diagnostic bundle](#)

CDP CLI commands for generating a diagnostic bundle

CDP includes a set of CLI commands that allow you to generate an on-demand diagnostic bundle that gathers logs and command output from your Data Lake, FreeIPA, or Data Hub cluster VMs. The bundle can also be used for support case troubleshooting if the logs captured via the equivalent UI option are not sufficient.

For each resource type (freeipa, datalake, or datahub), there are two commands that you can use:

- `get-<resource-type>-log-descriptors` - Allows you to obtain a list of logs that get collected for diagnostics.
- `collect-<resource-type>-diagnostics` - Triggers diagnostics collection. You need to provide a CRN of a resource (Data Lake, FreeIPA, or Data Hub cluster), a description, and a destination for the bundle.
- `list-<resource-type>-diagnostics` - Lists the running or finished diagnostic operations.
- `cancel-<resource-type>-diagnostics` - Stops the ongoing diagnostics operations.

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

Prerequisites

- 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 CM nodes.

`get-<resource-type>-log-descriptors`

The `get-<resource-type>-log-descriptors` commands allow you to obtain a list of logs that get collected. These commands can also be used to obtain available log label/path pairs that can be provided in the `collect-<resource-type>-diagnostics` command via the `--labels` parameter in order to limit the command output.

The output is printed in the Event History tab on the UI in Data Lake or Data Hub details. Since there is no FreeIPA Event History on the UI, in case of FreeIPA the output is saved to `/var/log/salt/minion`, which can be accessed from your logs location in cloud storage.

Commands

To gather available log label/path pairs, use one of the following commands:


```
cdp environments get-freeipa-log-descriptors
```

```
cdp datalake get-datalake-log-descriptors
```

```
cdp datahub get-datahub-log-descriptors
```

collect-<resource-type>-diagnostics

The collect-<resource-type>-diagnostics commands trigger log collection. You need to provide a CRN of a resource (Data Lake, FreeIPA, or Data Hub cluster), a description, and a destination for the bundle. The generated bundle can be saved to cloud storage or to a local file. The command progress (including the location where the bundle is generated) is printed in the Event History tab on the UI.



Note: This feature only works out of the box with environments created or upgraded in October 2020 or later. If your environment is older and your nodes have internet access, you can use the --update-package option to download the newest library that is needed for the collection.

Prerequisites

- If you would like to use cloud storage as a destination, ensure that the Logger and IDBroker instance profiles (on AWS) or managed identities (on Azure) are able to write to the specified cloud storage location.
- If you would like to use Cloudera support as a destination, your network must be able to reach the following destinations: <https://dbusapi.us-west-1.altus.cloudera.com> if your CDP account runs in Control Plane region us-west-1, or <https://api.<control-plane-region>.cdp.cloudera.com> if your CDP account runs in any other Control Plane region (only required if your cloud provider is AWS) and <https://cloudera-dbus-prod.s3.amazonaws.com> (required even if your cloud provider is not AWS).

Example commands

To run collection with salt script included and update cdp-telemetry package on the nodes, use one of the following commands, Replace <DESCRIPTION> with an actual description and the <NAME-OR-CRN>, <DATA LAKE-CRN>, and <DATAHUBCRN> with an actual CRN.

To send the diagnostics bundle to cloud storage, use:

```
cdp environments collect-freeipa-diagnostics --environment-name <NAME-OR-CRN> \
--description <DESCRIPTION> \
--destination CLOUD_STORAGE \
--include-salt-logs
```

```
cdp datalake collect-datalake-diagnostics --crn <DATA LAKE-CRN> \
--description <DESCRIPTION> \
--destination CLOUD_STORAGE \
--include-salt-logs
```

```
cdp datahub collect-datahub-diagnostics --crn <DATAHUB-CRN> \
--description <DESCRIPTION> \
--destination CLOUD_STORAGE \
--include-salt-logs
```

To send the diagnostic bundle directly to Cloudera support, use:

```
cdp datahub collect-datahub-diagnostics --crn <DATAHUB-CRN> \
--description <DESCRIPTION> \
--case-number <SUPPORT_CASE_NUMBER> \
--destination SUPPORT \
--update-package \
--include-salt-logs
```

To collect only specific custom logs (such as `/var/log/audit/audit.log` used in this example), use:

```
cdp environments collect-freeipa-diagnostics --environment-name <name> \
--description <DESCRIPTION>
--destination CLOUD_STORAGE \
--additional-logs path=/var/log/audit/audit.log,label=audit \
--labels audit
```

If you run this command:

- Logs specified under `--additional-logs` will be collected
- Due to the specified label, all other logs will be filtered out and no other logs will be included in the bundle.

Required parameters

`--environment-name` (for Data Lake or Data Hub) or `--crn` (for Data Hub) Provide a CRN of your Data Lake, FreeIPA, or Data Hub cluster. You can obtain it by running a `list` or `describe` command on your resource.

`--description` Provide a description for the diagnostics bundle.

`--destination` Specify a destination of the diagnostics collection: `SUPPORT`, `CLOUD_STORAGE`, or `LOCAL`:

- `SUPPORT`: Sends diagnostic bundles directly to Cloudera support.
- `CLOUD_STORAGE`: Saves collected diagnostics into a `tar.gz` file and uploads it to your logs cloud storage location specified during environment registration, to a `PATH` that looks like this:

```
cluster-logs/<CLUSTER-TYPE>/<CLUSTER-NAME>_<UUID PART-OF-CLUSTER-CRN>/d
iagnostics.
```

The “cluster logs” directory is only created if your cloud storage location is a bucket. If your cloud storage location includes a directory within a bucket, then the “cluster logs” directory is not created.



Note: The value of `<CLUSTER_NAME>` for FreeIPA is `<ENVIRONMENT-NAME>-freeipa`. For example, if your environment is called “test”, the `<CLUSTER_NAME>` for FreeIPA is “test-freeipa”.

- `LOCAL`: Diagnostics will be collected to a compressed file, and saved to `/var/lib/filecollector` on every cluster node.

Optional parameters

`--labels` You can use this to pass a list of labels that can be used to filter collected logs. This is useful if instead of including all logs, you would only include certain specific logs in the bundle. This list can be obtained from the `get-<cluster-type>-log-descriptors` response.

`--additional-logs` One or more additional VM log objects that will be included in the diagnostics bundle. A VM log object is a `path/label` pair. Cloudera support may ask you to provide specific values here.

`--hosts` Run diagnostics only the specified hosts (both IP addresses or fqdns are valid).

`--host-groups` Run diagnostics only on the those hosts that are included in the specified host groups.

`--exclude-hosts` Same as the `--hosts` filter but for exclusion.

`--skip-unresponsive-hosts` Skip unhealthy hosts from the diagnostics collection (where “unhealthy” means that salt minion is unresponsive). If this is not used, the diagnostics will fail if there is an unresponsive host during diagnostics.

`--include-salt-logs` If this is set, salt minion/master logs will be included in the diagnostic bundle.

`--case-number` Allows you to provide a support case number.

`--update-package` If this is set, during diagnostics collection, the required third party applications will be downloaded and updated on the VM nodes. This requires internet access.

`--storage-validation` You can use this if using `CLOUD_STORAGE` as a destination. If this is set, cloud storage write validation check will happen during an initial phase of the diagnostics collection.

For more information about the command and its options, use the `--help` CLI option.

list-<resource-type>-diagnostics

Lists the running or finished diagnostic operations.

Examples:

```
cdp environments list-freeipa-diagnostics --environment-name <NAME-OR-CRN>
cdp datalake list-datalake-diagnostics --crn <DATA LAKE-CRN>
cdp datahub list-datahub-diagnostics --crn <DATA HUB-CRN>
```

cancel-<resource-type>-diagnostics

Stops the ongoing diagnostics operations.

Examples:

```
cdp environments cancel-freeipa-diagnostics --environment-name <NAME-OR-CRN>
cdp datalake cancel-datalake-diagnostics --crn <DATA LAKE-CRN>
cdp datahub cancel-datahub-diagnostics --crn <DATA HUB-CRN>
```

Related Information

[Diagnostic bundle content](#)

[Send a diagnostic bundle to Cloudera Support](#)

[Downloading Impala diagnostic bundles from AWS](#)

[Downloading Diagnostic Bundles for a Workspace](#)

Diagnostic bundle content

This topic summarizes the content of a diagnostic bundle.

The following information is included in a diagnostic bundle:

- Cluster details required to identify the bundle (account ID, cluster CRN, cluster type, and so on)
- Redacted service and system logs
- Web server configuration from the salt master node (nginx)
- An HTML report including:
 - The result of running a number of system-level commands on each host
 - Network endpoint accessibility for outbound network traffic requirements
 - Health details about the required infra services
- When triggering diagnostic bundle collection via CDP CLI, you may specify additional logs that are not part of the default log collection to be collected in the diagnostic bundle.

The information is collected from all running nodes of your cluster (Data Lake, FreeIPA, or Data Hub). It is not collected from nodes that are not running, such as nodes that are currently being provisioned or stopped or deleted nodes.



Note: The logs collected by default in the on-demand diagnostic bundle are the same as when you [enable cluster log collection](#). Whether or not you enabled the log collection feature on tenant or environment level, the on-demand diagnostic bundle can be generated independently when needed.



Note: The Cloudera Manager service logs can also be gathered via Cloudera Manager, as described in [Collecting Data Lake diagnostics via CM](#). However, if Cloudera Manager is down or cannot be started, generating a diagnostic bundle may help troubleshoot CM issues.