

Troubleshooting Cloudera Data Engineering

Date published: 2020-07-30

Date modified: 2023-11-02



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

Cloudera Data Engineering log files.....	4
Viewing Job run timeline.....	4
Using Spark history server to troubleshoot Spark jobs.....	6
Monitoring Data Engineering service resources with Grafana	
dashboards.....	6
Connecting to Grafana dashboards in Cloudera Data Engineering Private Cloud.....	7
Accessing Grafana dashboards.....	10
Downloading diagnostic bundles for CDE.....	11

Cloudera Data Engineering log files

You can view logs for Cloudera Data Engineering (CDE) using the web console, including CDE service logs, virtual cluster logs, and job logs.

To view logs for a CDE service or virtual cluster, click the three-dot menu for the service or virtual cluster, and then select View Logs. When the View Logs modal is displayed, you can download the logs or copy them to the clipboard by clicking the associated icon at the top right of the modal.

To view logs for a job run:

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Locate the Virtual Cluster containing the job that you want to troubleshoot.
3. Click View Jobs for the cluster containing the job.
4. Select the job that you want to troubleshoot.
5. In the Run History tab, click the Run ID for the job run that you want to troubleshoot.
6. Go to the Logs tab.
7. Select the log you want to view using the Select log type drop-down menu and the log file tabs.
8. To download the logs, click the Download menu button. You can download a text file of the currently displayed log, or download a zip file containing all log files.

Viewing Job run timeline

You can view the intermediate stages of the job run at every stage during its life cycle in real-time.

About this task

In case of a job failure, you can view the specific event and component where the job run failed. This reduces turnaround time during the debugging process for job run failure. You can see the step-by-step advancement of the job run on the UI, including all the granular details instead of reviewing extensive logs to obtain the same insights.

Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. In the left navigation pane, click Jobs Runs. The Jobs Runs page displays.
3. Click on the Job Id for which you want to see the status.

4. Go to the Timeline tab. It displays the summary of the Job run progression in the reverse chronological order for both primary and subordinate stages.

The screenshot displays the Cloudera Data Engineering web interface. On the left is a dark sidebar with navigation links: Home, Jobs, Job Runs (highlighted with an orange box), Sessions, Repositories, Resources, and Administration. The main content area on the right shows the breadcrumb 'cluster-01 / Job Runs /' and a 'Status' section indicating 'Succeeded'. Below this, the 'Timeline' tab is selected and highlighted with an orange box, with a 'New' button next to it. The 'Summary' section, titled 'View your Job run progression i', lists three job stages: 'Succeeded' (Component: Runtime API Server), 'Running' (Component: Spark Container), and 'Starting' (Component: Runtime API Server). Each stage includes a 'Show Details' link.

Using Spark history server to troubleshoot Spark jobs

The Spark history server is a monitoring tool that displays information about completed Spark applications. It provides information for debugging such as Spark configurations, DAG execution, driver and executor resource utilization, application logs, and job, stage and task-level details.



Important: To access and load the Spark UI correctly, ensure that the third party cookies for `[*].cloudera.com` and `[*].cloudera.site` are enabled.

To view spark history server information for a job run:

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Click Jobs in the left navigation menu.
3. From the drop-down in the upper left-hand corner, select the Virtual Cluster that you want to restore jobs to.
4. Select the job that you want to troubleshoot.
5. Click Jobs Runs in the left menu, and click the Run ID for the job run you want to view the information.
6. Click the Spark UI tab to access the Spark History Server.

Monitoring Data Engineering service resources with Grafana dashboards

Grafana is a visualisation and analytics software that enables the development of dashboards to monitor metrics data. You can access pre-built Grafana dashboards to monitor your jobs and virtual clusters in Cloudera Data Engineering (CDE).

The CDP metrics are stored centrally in the Prometheus database and monitored by Prometheus. Grafana uses these metrics for data visualization. Your workload databases are not involved in any way.

You can immediately view the following pre-built dashboards for viewing runtime metrics in CDE:

Kubernetes Dashboard

This dashboard includes generalized visualizations of CDE job run statuses. It displays the following information:

- Number of succeeded, failed, and killed jobs for the given period
- Total number of jobs in the Starting phase
- Total number of jobs in the Running phase

Virtual Cluster Metrics Dashboard

This dashboard includes visualizations of service requests, pod counts and job statuses for the selected Virtual Cluster. The available metrics are:

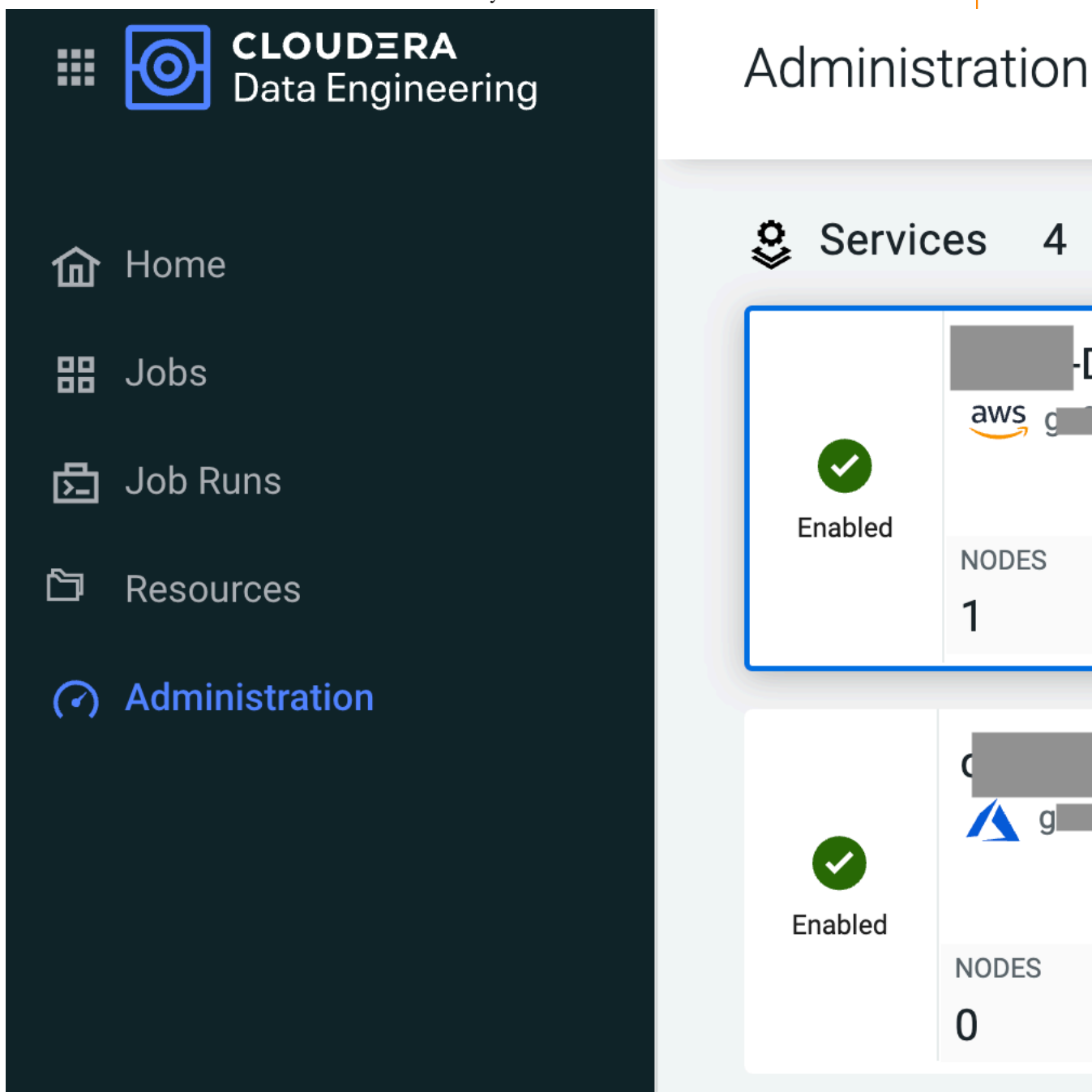
- Time series of CPU requests of running pods (includes virtual cluster service overhead)
- Time series of memory requests of running pods (includes virtual cluster service overhead)
- The response time of Livy's requests
- Time series for the number of pods in running and pending states
- Total number of running pods and pending pods
- Time series of starting and running jobs, and the total number of successful jobs

Connecting to Grafana dashboards in Cloudera Data Engineering Private Cloud

This topic describes how to access Grafana dashboards for advanced visualization of Virtual Cluster's metrics such as memory and CPU usage in Cloudera Data Engineering (CDE) Private Cloud.


For CDE Service

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The Home page displays.
2. Click Administration in the left navigation menu and locate a Service in the Services column and click Service Details on the environment for which you want to see the Grafana dashboard.



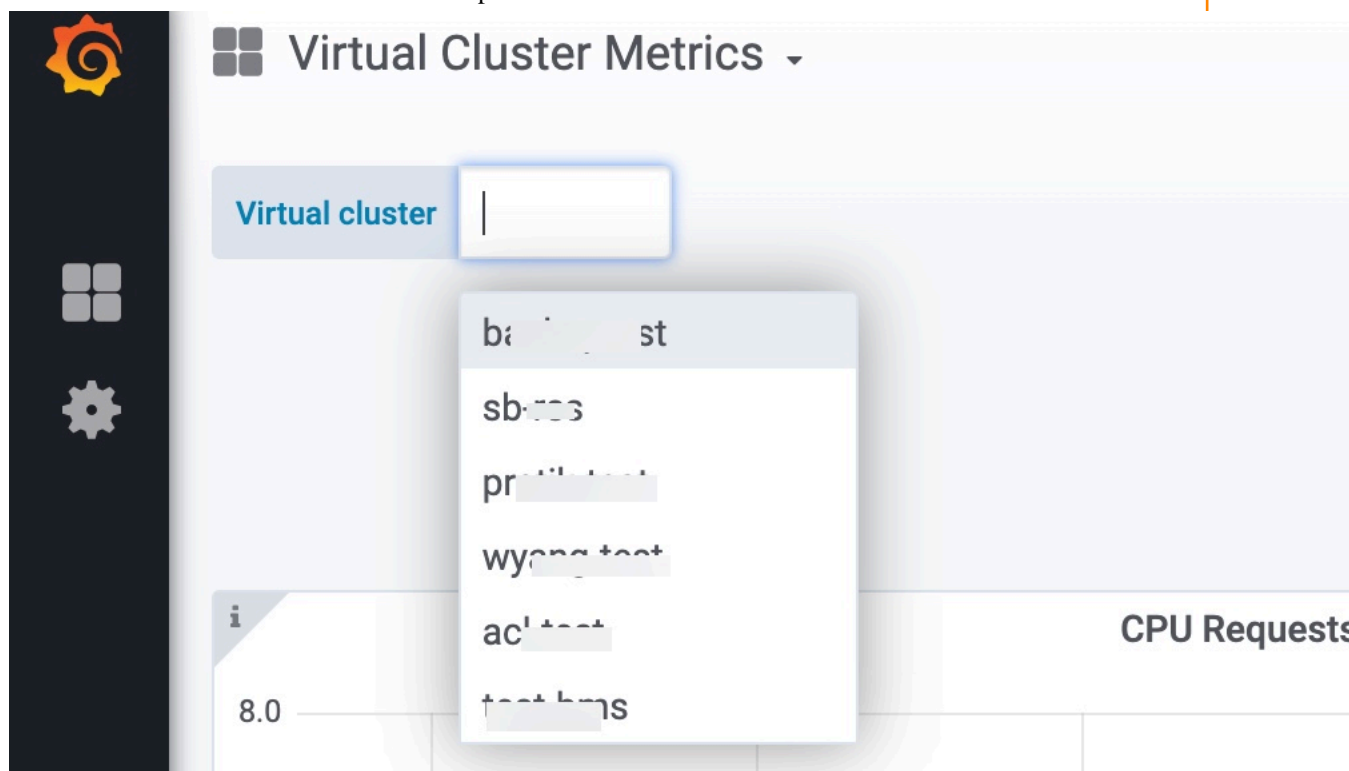
3. In the Administration/Service page, click Grafana Charts. A read-only version of the Grafana interface opens in a new tab in your browser.

4.

In the Grafana dashboard, click the grid icon  in the left navigation menu.

5. Select Virtual Cluster Metrics under the Dashboards pane.

6. Click on a virtual cluster name from the dropdown list to view the Grafana charts.



about CPU requests, memory requests, jobs, and other information related to the virtual cluster is displayed.

For Virtual Cluster


1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. In the Virtual Clusters section, navigate to the virtual cluster for which you want to see the Grafana dashboard.
3. Click View Cluster Details for the virtual cluster.

The Administration/Virtual Cluster page is displayed.

4. Click Grafana Charts.



A read-only version of the Grafana interface opens in a new tab in your browser.

Overview / [s...](#)

 **Running**

[s...](#) **23Mar**

VERSION	VC ID	CREATED BY	CPU	MEMORY	JOBS
1.1...	-b26	dex-ann-600+77va	S...	0	0 B
					0 🔗

[CLI TOOL](#) : [API DOC](#)  [JOBS API URL](#)  **[GRAFANA CHARTS](#)**

[Configuration](#) [Charts](#) [Logs](#)

CDE Service

[c...](#)

Information about CPU requests, memory requests, jobs, and other information related to the virtual cluster is displayed.

- 5.** In the Virtual Cluster Metrics page, click on a virtual cluster name from the Virtual Cluster dropdown list to view the Grafana charts of that virtual cluster.



Note: You must first view the charts using the GRAFANA CHARTS link. Only then the charts in the Charts tab get loaded. Otherwise, it displays the The web page at <https://service.cde-nrjcrwg7.apps.apps.shared-01.kcloud.cldr.com/grafana/d/usZz/kubernetes?kiosk> might be temporarily down or it may have moved permanently to a new web address. error.




Accessing Grafana dashboards

Cloudera provides pre-built Grafana dashboards comprising metrics data, charts, and other visuals. You can access pre-built Grafana dashboards to monitor your jobs and virtual clusters in Cloudera Data Engineering (CDE). You can immediately view the Kubernetes and Virtual Cluster Metrics pre-built dashboards in CDE.

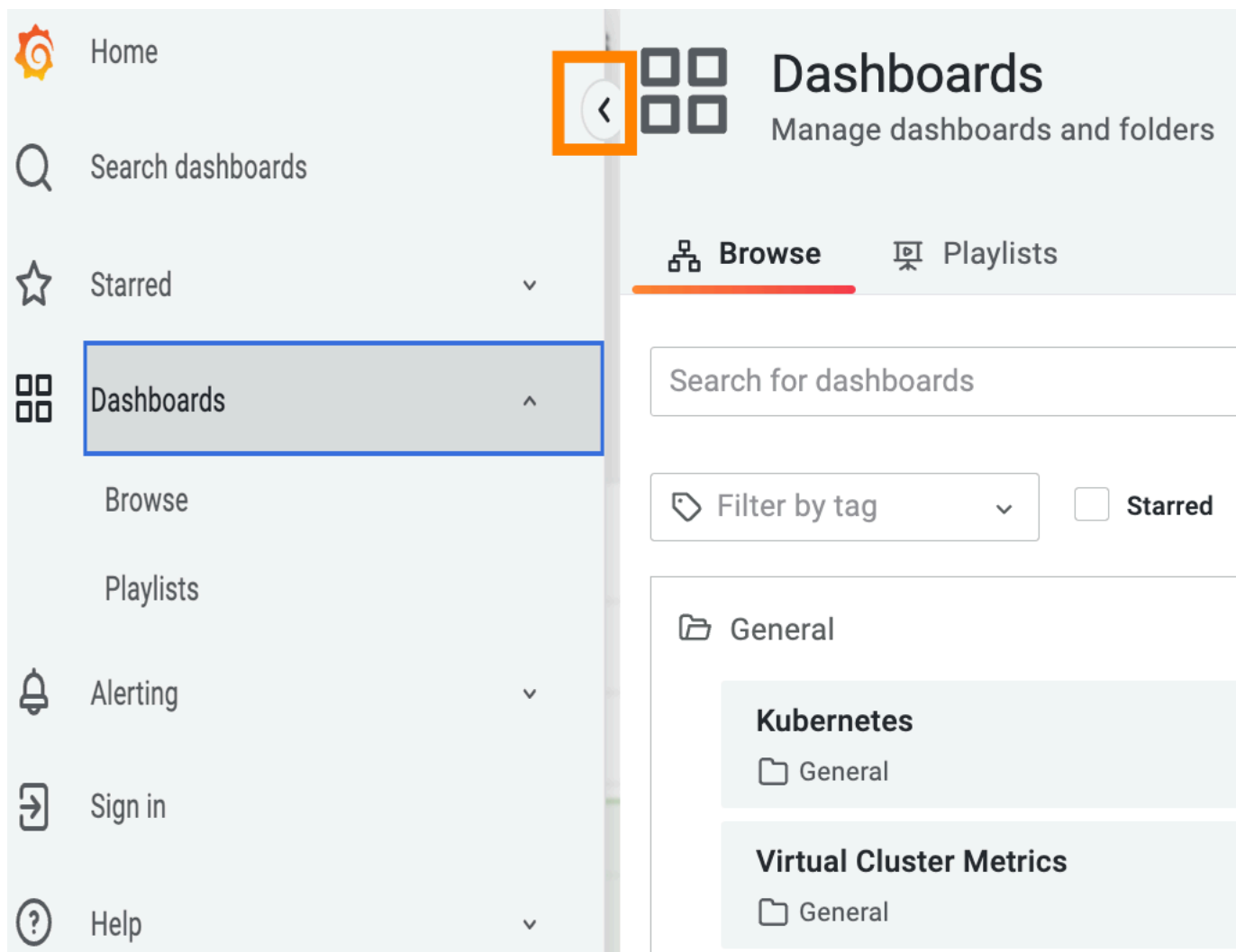
Before you begin

You must first connect to the Grafana dashboards in CDE Private Cloud to view the Kubernetes and Virtual Cluster Metrics dashboards.

Procedure

1. After you connect to the Grafana Dashboards from the CDE UI, click the  icon to view the left navigation pane.
2. Click Dashboards > Browse. The Dashboards screen is displayed.

3. In the Browse tab of Dashboards, click Kubernetes or Virtual Cluster Metrics to view the respective dashboard.



Downloading diagnostic bundles for CDE

Learn how to manage and download diagnostic bundles.

The CDP platform provides various services for managing and downloading diagnostic bundles.

You can download diagnostic bundles from the Cloudera Data Engineering. For more information, see *Options for generating the CDP Private Cloud diagnostic data*.

You can also send usage and diagnostic data from Cloudera Manager. For more information, see *Sending Usage and Diagnostic Data to Cloudera*.

Related Information

[Options for generating the CDP Private Cloud diagnostic data](#)

[Sending Usage and Diagnostic Data to Cloudera](#)