Cloudera Data Engineering 1.4.1

Troubleshooting Cloudera Data Engineering

Date published: 2020-07-30
Date modified: 2022-11-18



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. 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

Using Spark history server to troubleshoot Spark jobs.... 4

Connecting to Grafana dashboards in Cloudera Data Engineering Private Cloud.... 4

Cloudera Data Engineering diagnostic bundles and summary status....8

Downloading summary status for Cloudera Data Engineering.... 8

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. From the CDE Overview page, select the CDE service for the job you want to troubleshoot.
2. In the Virtual Clusters column, click the View Jobs icon for the cluster containing the job.
3. Select the job you want to troubleshoot.
4. In the Run History tab, click the Job ID for the job run you want to troubleshoot.
5. Go to the Logs tab.
6. Select the log you want to view using the Select log type drop-down menu and the log file tabs.
7. 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.

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.

To view spark history server information for a job run:

1. From the CDE Overview page, select the CDE service for the job you want to troubleshoot.
2. In the Virtual Clusters column, click View Jobs for the cluster containing the job.
3. Select the job you want to troubleshoot.
4. In the Jobs Run page, click the Job ID for the job run you want to view the information.
5. Click the Spark UI tab to access the Spark History Server.

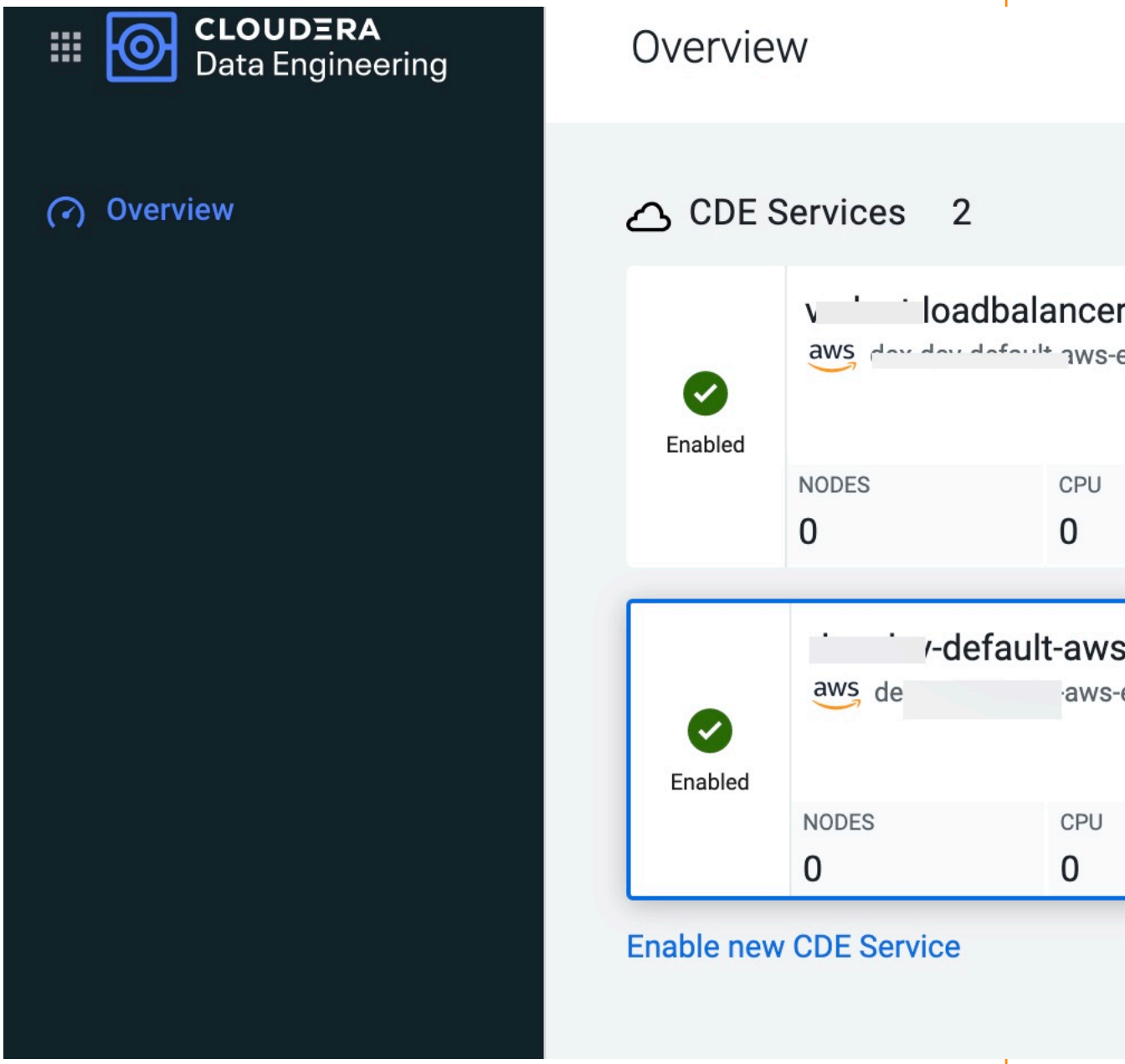
Connecting to Grafana dashboards in Cloudera Data Engineering Private Cloud

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

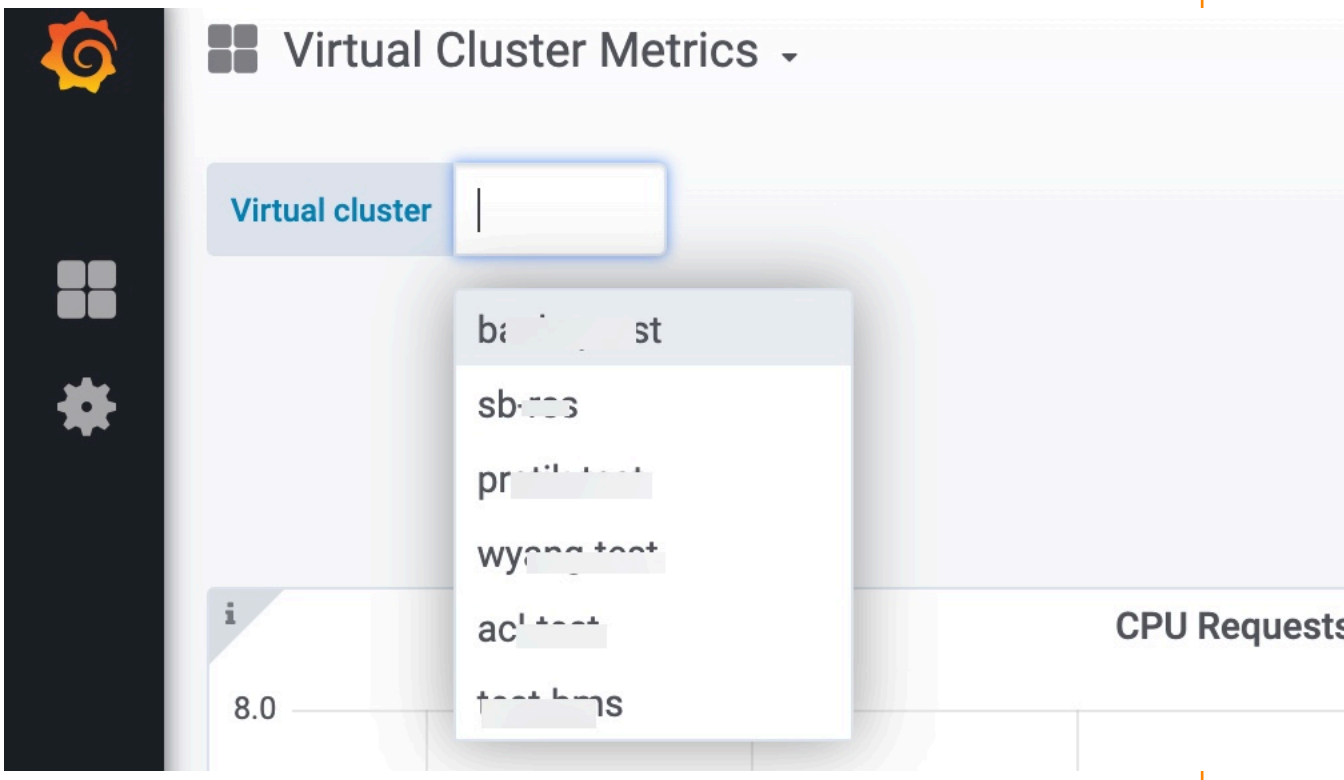
1. Navigate to the Cloudera Data Engineering Overview page by clicking the Data Engineering tile in the Cloudera Data Platform (CDP) management console.

2. In the CDE Services column, click the Service Detailsbutton on the environment for which you want to see the Grafana dashboard.



3. In the Environment 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.

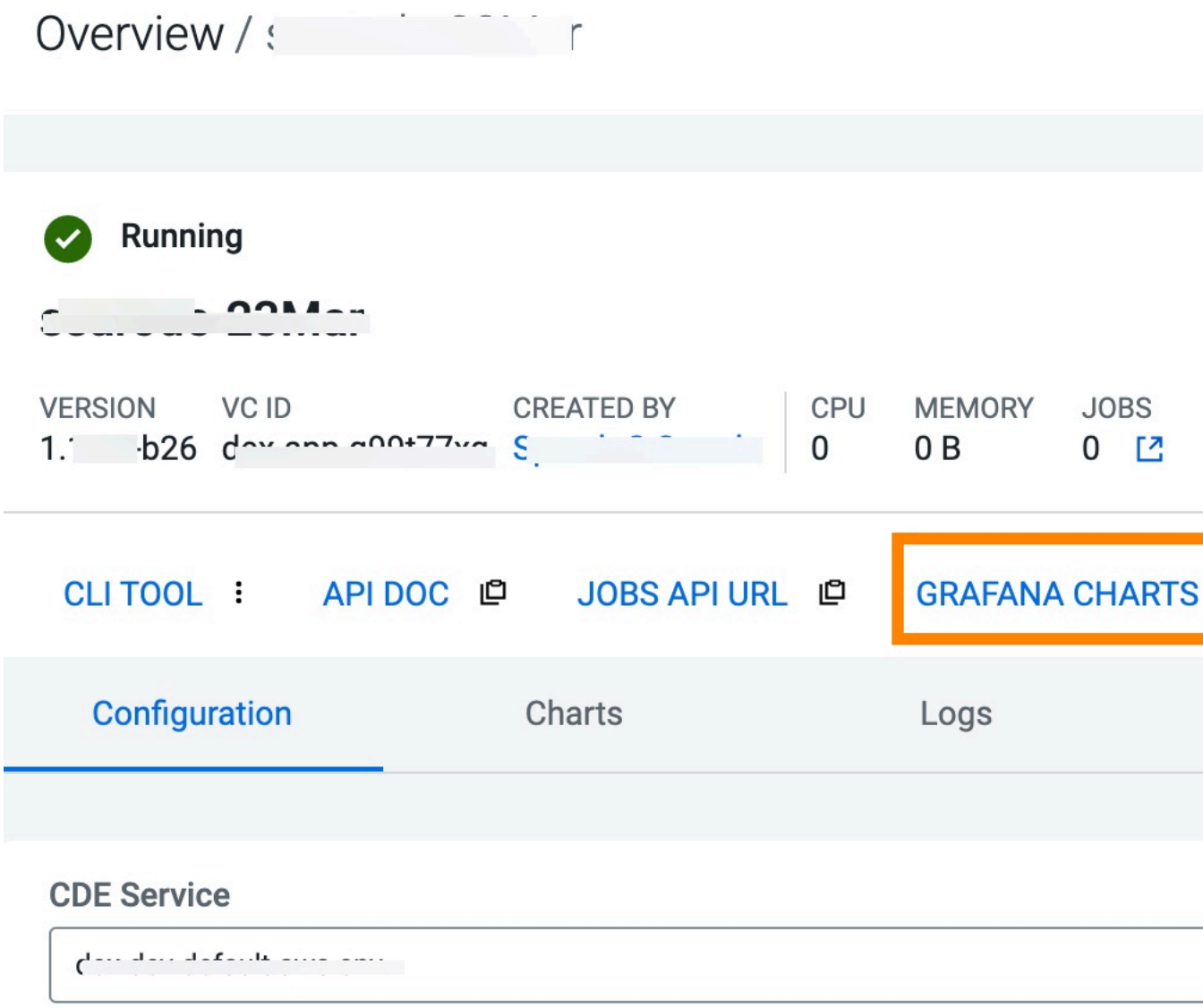
For Virtual Cluster

1. Navigate to the Cloudera Data Engineering Overview page by clicking the Data Engineering tile in the Cloudera Data Platform (CDP) management console.
2. In the Environment column, select the environment containing the virtual cluster for which you want to see the Grafana dashboard.
3. In the Virtual Clusters column on the right, click the Cluster Details icon of the virtual cluster.

   The virtual cluster's Overview page is displayed.

4. In the Overview page, click Grafana Charts.

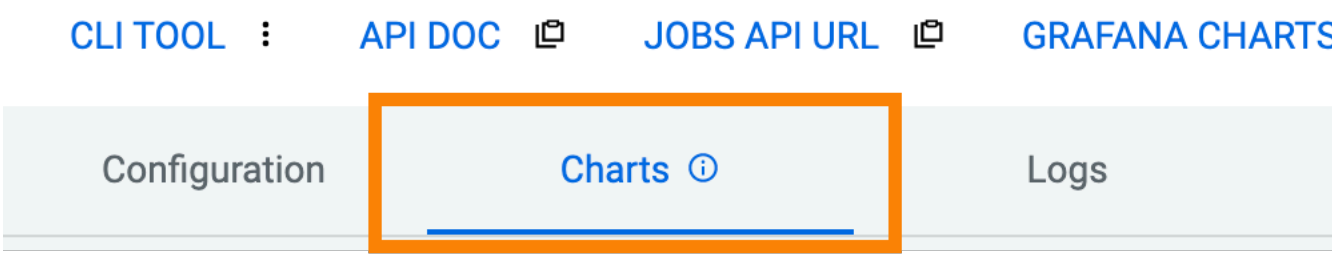
   A read-only version of the Grafana interface opens in a new tab in your browser.



   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.



Cloudera Data Engineering diagnostic bundles and summary status

Cloudera Data Engineering provides the ability to download log files, diagnostic data, and a summary status for the running CDE services and virtual clusters. You can provide this data to Cloudera Support for assistance troubleshooting an issue.

The following section describes about CDE Diagnostic Bundles and Summary Status, and the information collected in each.

Diagnostic Bundle

Important: The CDE service inCDP Private Cloud Data Services does not currently support diagnostic bundles.

The diagnostic bundle is a collection of the logs from the CDE Services. These logs can be downloaded on-demand from CDE UI. CDE gives you the functionality to select the sources for which you want to download the logs and you can also select a predefined or custom time range for these logs.

Information Collected in Diagnostic Bundles

- Container logs for all running CDE service pods (excluding user compute pods).

Summary Status

The Status Summary shows the status of each service instance being managed by the CDE Control Plane. It is a package of JSON files consisting of information related to configuration, monitoring, logging, events and health test results of the service and its instances.

Information Collected in Summary Status

- Status of all cloud resources created during CDE provisioning
- Kubernetes resource status for critical infrastructure pods, deployments, pods, services and events for core service infrastructure pods or virtual cluster infrastructure pods

Downloading summary status for Cloudera Data Engineering

This section describes how to download a summary status of a Cloudera Data Engineering (CDE) Service in Cloudera Data Platform (CDP) using CDE UI.

About this task

To get a current snapshot of the current CDE infrastructure status (cloud resources, Kubernetes pod statuses, pod events and service metadata). This summary status, in the form of a ZIP file, is downloaded to your computer. This is available on both AWS and Azure environments.

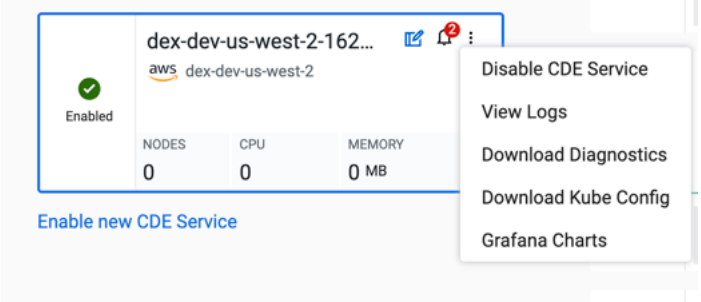
Before you begin

Required Role: DEAdmin

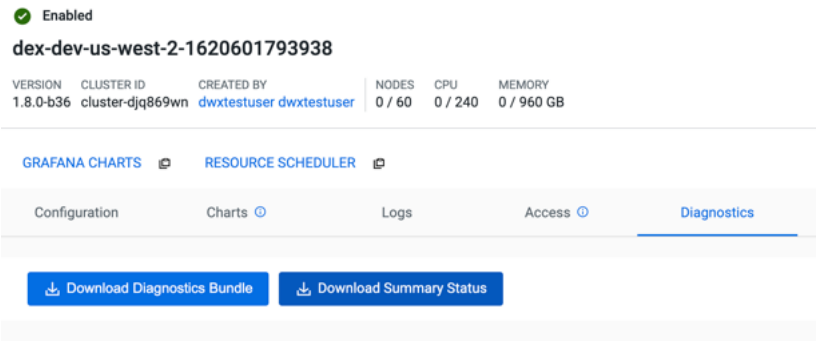
After granting or revoking the role on the environment, run the Sync Users to FreeIPA environment action.

Procedure

1. Log in to the CDP web interface and navigate to the CDE service.
2. In the CDE service, click Overview in the left navigation panel, and select the CDE Service for which you want to download a summary status.
3. In the selected CDE service, click on the three dots to see a drop down menu, and select the Download Diagnostics option.



4. It takes you to the Diagnostics tab of the Service details. Click the Download Summary Status button.



5. Summary Status gets downloaded on your local machine in a Zip file.

Results

When you extract the Summary Status ZIP file that you downloaded from CDE UI, you can see the directories that contain files in JSON format.

Below is an example of format of the Zip file for Summary Status:

Note: In the extracted Summary Status file, cluster.json refers to the CDE service and ‘instances’ refer to CDE virtual clusters.

```
cde-service-diagnostics-{clusterID}-{timestamp}/
  cluster.json            # output from cluster describe
  cluster-events.json     # output from get cluster events
  cloud-diagnostics.json  # cloud resource status (RDS, EFS, EKS, LB, ...)
```

```
  instances/
    {id}.json             # output from instance id describe
    {id}-events.json      # output from get instance id events
  kubernetes/             # kubernetes resource status for all relevant obj
ects and namespaces
    nodes.json
    {namespace}/
      namespace.json
      pods.json
      events.json
      deployments.json
      replicasets.json
      services.json
      daemonsets.json
    {namespace}/
      namespace.json
      pods.json
      events.json
      deployments.json
      replicasets.json
      services.json
      daemonsets.json
    ...
```