Cloudera Data Warehouse Private Cloud 1.5.0

Troubleshooting

Date published: 2020-08-17
Date modified: 2023-01-25

https://docs.cloudera.com/

Legal Notice

Contents

Locating Cloudera Data Warehouse Private Cloud logs

Learn how you can access logs for Cloudera Data Warehouse (CDW) Private Cloud.

About this task

When you generate logs using the Collect Diagnostic Bundles option from the environment, they are written to a partition on the Hive sys.logs table and are stored in the following location on HDFS:

```
/warehouse/[***ENVIRONMENT-NAMEabcde***]/[***DATABASE-CATALOG***]/warehouse/
tablespace/external/hive/sys.db/logs
```

“abcde” is a random 5-character string that is appended to the environment name.

These partitions are retained for 7 days by default.

Procedure

1. Log in to the OpenShift or Experiences Compute Service (ECS) cluster and determine the location of the sys.logs table by running the following query:

   ```
   DESCRIBE FORMATTED sys.logs;
   ```

   This SQL statement returns information about the location of the table which contains the logs.
2. Use the location obtained in Step 1 to locate the CDW Private Cloud logs on the OpenShift or ECS clusters.

Downloading Hive diagnostic bundles in Data Warehouse Private Cloud

You can download diagnostic bundles for troubleshooting a Hive Virtual Warehouse in Cloudera Data Warehouse (CDW) Private Cloud. The diagnostic bundles contain log files for the sidecar containers that support Hive components and for the components themselves. These diagnostic bundles are stored on HDFS in the form of ZIP files.

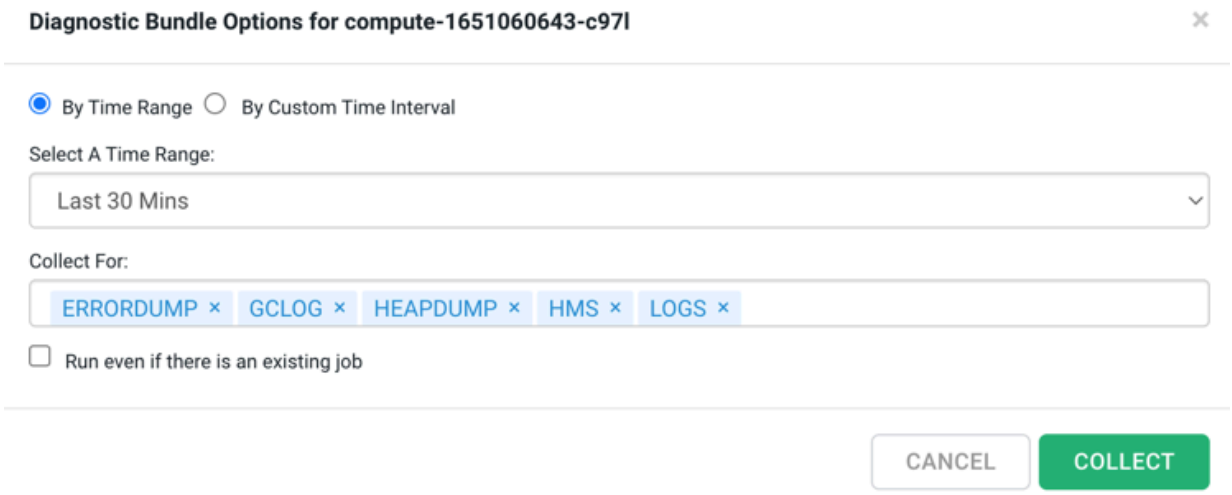
About this task

The log files are generated when you run some workloads on your Hive Virtual Warehouse.

Procedure

1. Log in to the Data Warehouse service as a DWAdmin.

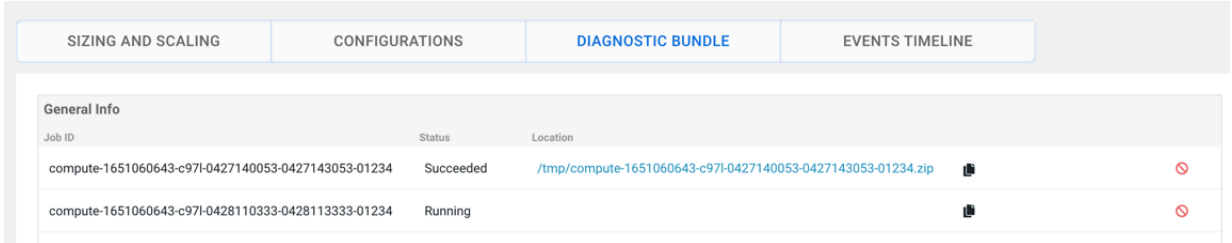
2. Go to a Hive Virtual Warehouse and click ⋮ Collect Diagnostic Bundle .
   The options for generating the diagnostic bundles are displayed as shown in the following image:



3. Select the time period for which you want to generate the logs.
   - Select the By Time Range option to generate logs from last 30 minutes, one hour, 12 hours, or 24 hours.
   - Select By Custom Time Interval option to generate logs for a specific time period based on your requirement.

   Note: You must set the time range as per the UTC timezone.

4. Select the categories for which you want to generate the logs by selecting the options from the Collect For section.
   By default, ERRORDUMP, GCLOG, HEAPDUMP, HMS, LOGS, CRINFO, K8S-RESOURCE-INFO are selected. Click X to remove the ones you do not need.
   - ERRORDUMP contains exceptions from the containers
   - CGLOG contains JVM garbage collector-related logs
   - HEAPDUMP contains JVM heapdump
   - HMS contains sidecar container logs that support the metastore
   - LOGS contains logs of Hive, Coordinator, and Executor processes and their supporting containers
5. Select the Run even if there is an existing job option to trigger another diagnostic bundle creation when one job is running.
6. Click Collect.
   The following message is displayed: Collection of Diagnostic Bundle for compute-1651060643-c97l initiated. Please go to details page for more information.
7. Go to the Virtual Warehouses details page by clicking ⋮ Edit .
8. Go to the **DIAGNOSTIC BUNDLE** tab.
   The jobs that have been triggered for generating the diagnostic bundles are displayed, as shown in the following image:



9. Click on the link in the Location column to download the diagnostic bundle to your computer.

Impala queries fail

Condition

Impala queries running with high concurrency fail on Embedded Container Service (ECS) with the following errors: Invalid or unknown query handle and Invalid session id.

Cause

Impala queries might fail because a single ECS server may not be able to handle the load. To resolve this issue, enable ECS High Availability and increase the ECS server replicas. This process is called promoting the ECS agents to servers. You must promote only one ECS agent at a time. This procedure is explained using an example where you promote the ECS agent on agent1.example.com and then promote the ECS agent on agent2.example.com.

Solution

Procedure

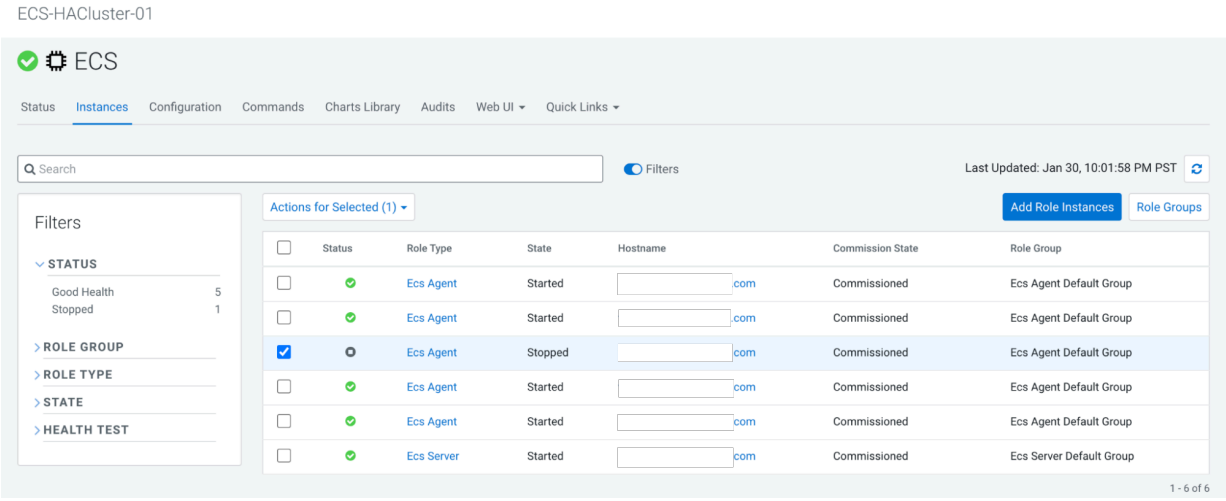
1. Prepare the agent node for promotion by running the following commands on the command line of your ECS server host.

```
sudo /var/lib/rancher/rke2/bin/kubectl --kubeconfig=/etc/rancher/rke2/rke2.yaml get nodes
```

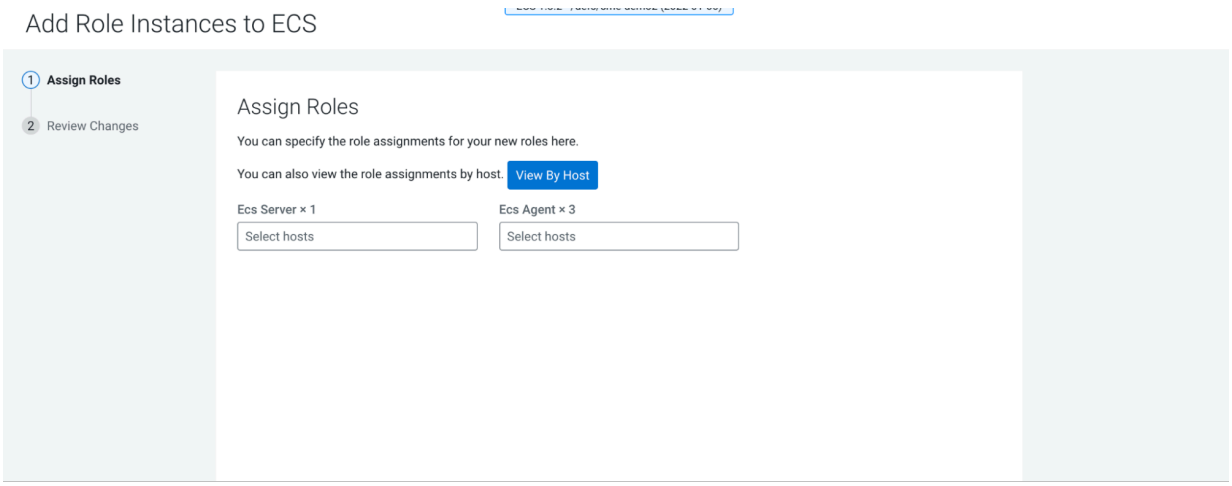
```
sudo /var/lib/rancher/rke2/bin/kubectl --kubeconfig=/etc/rancher/rke2/rke2.yaml drain agent1.example.com --ignore-daemonsets --delete-emptydir-data
```

Note: This may take a few minutes.

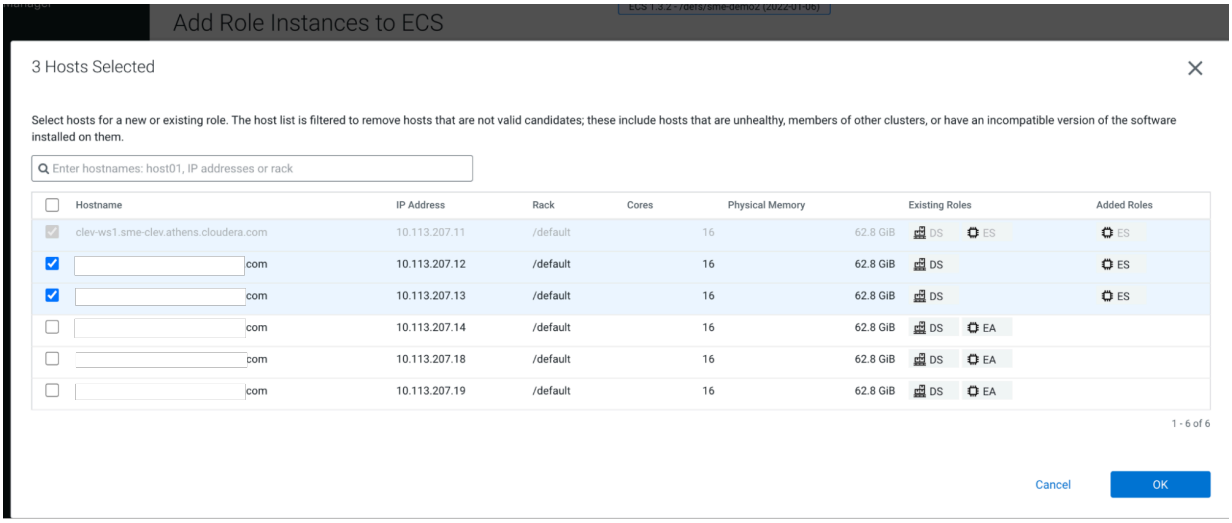
2. In Cloudera Manager, navigate to ECS Cluster ECS . Stop the ECS Agent running on agent1 and then delete the agent by selecting the respective option from the Actions for Selected drop-down menu.

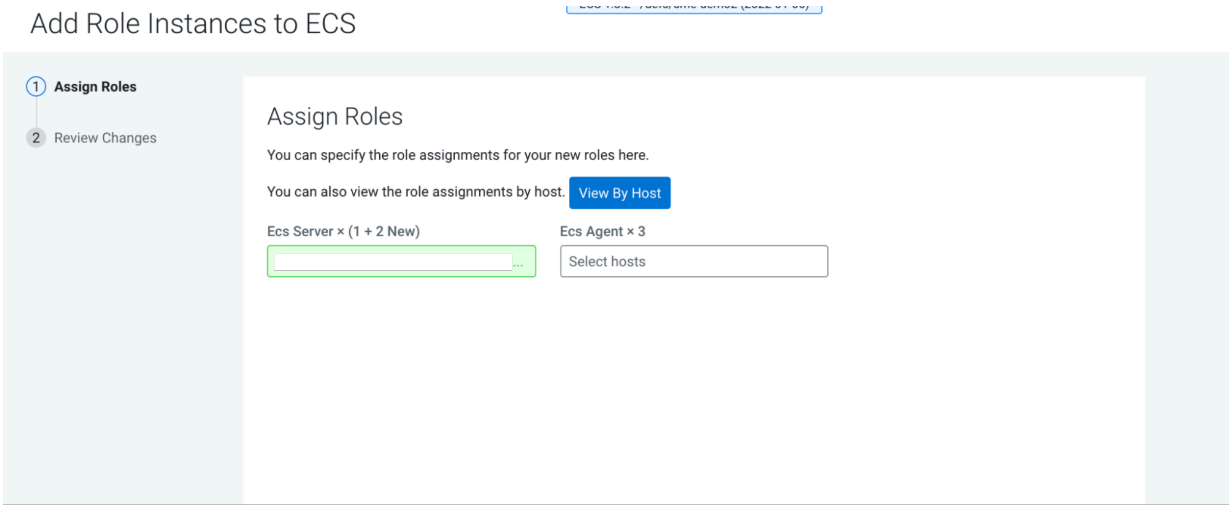


3. In Cloudera Manager, navigate to ECS Cluster ECS and click Add Role Instances.

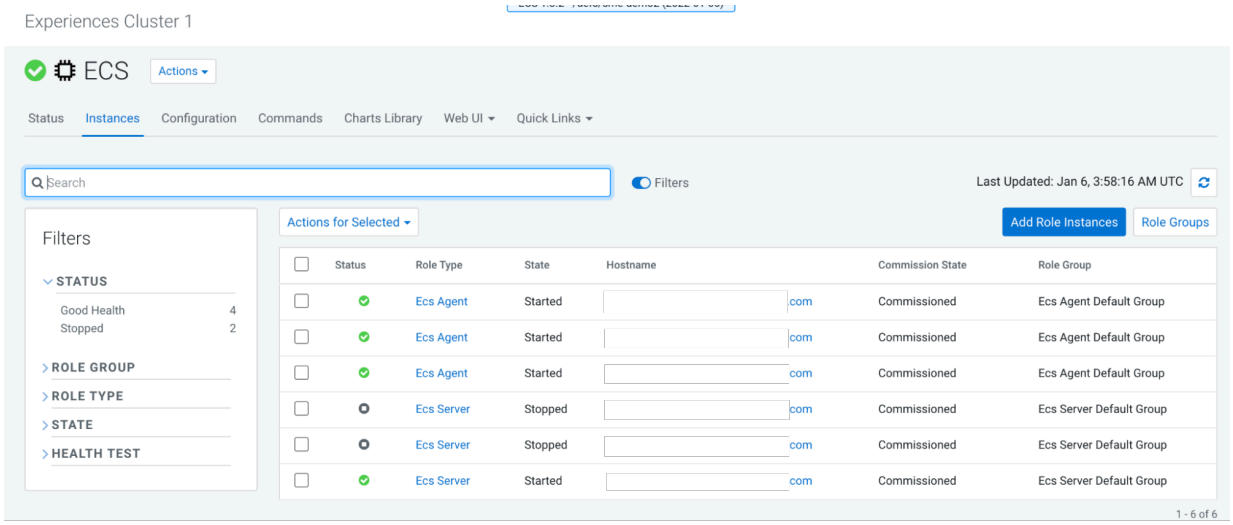


4. Add the available host agent1 as an ECS server in the Add Role Instances to ECS pop-up. Click Ok.

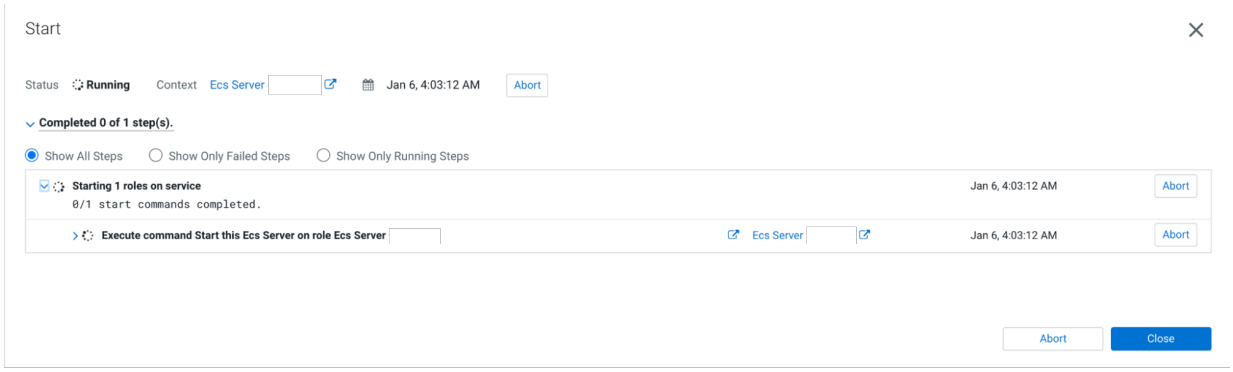




5. Click Continue.



6. Start the new ECS server from ECS Instances view. For example, start ECSServer on agent1.



7. On the command line, uncordon the node by running the following command:

```
sudo /var/lib/rancher/rke2/bin/kubectl --kubeconfig=/etc/rancher/rke2/rke2.yaml uncordon agent1.example.com
```

8. Confirm the node's status from webUI or the command line by running the following command:

```
sudo /var/lib/rancher/rke2/bin/kubectl --kubeconfig=/etc/rancher/rke2/rke2.yaml get nodes
```

Note: Do not proceed until node status is Ready. This may take several minutes.