Cloudera Data Warehouse Private Cloud 1.4.1

CDW Private Cloud Administration

Date published: 2022-11-18
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

Enabling Data Analytics Studio....4

Disabling Data Analytics Studio....4

About predefined Kerberos principals....4

About the Refresh option....5

Third-party object storage support....5

Using Ozone storage with Cloudera Data Warehouse Private Cloud....6

Creating a database on Ozone for Cloudera Data Warehouse Private Cloud Virtual Warehouses....6
Configuring Hive/Impala logging on Ozone for Cloudera Data Warehouse Private Cloud....7
Specify or create an Ozone bucket for Cloudera Data Warehouse Private Cloud logs....8
Update Cloudera Data Warehouse Private Cloud log configuration to point to Ozone....8
Monitor Cloudera Data Warehouse Private Cloud logs on Ozone storage....10
Analyze Cloudera Data Warehouse Private Cloud logs stored on Ozone....11

Changing delegation username and password....11

SSL-enabled endpoints for Virtual Warehouse clients in Cloudera Data Warehouse Private Cloud....12

Debugging with Impala Web UIs....13

Generating and downloading diagnostic bundles....14

Configuring Impala Virtual Warehouses to encrypt spilled data in Cloudera Data Warehouse Private Cloud....16

Customizing Impala pod configuration....16

Enabling Data Analytics Studio

Data Analytics Studio (DAS) has been deprecated in CDW Private Cloud 1.4.1 and higher and is disabled by default. Cloudera encourages you to use Hue for Hive and Impala workloads. However, if you still need DAS, you can enable it from the Advanced Configuration.

Procedure

1. Log in to the Data Warehouse service as an Administrator.
2. Click Advanced Configuration to go to the **Advanced Settings** page.
3. Select the Enable DAS option.
4. Click Update.

What to do next

Create a new Database Catalog and a Hive Virtual Warehouse.

Disabling Data Analytics Studio

Data Analytics Studio (DAS) in Data Warehouse Private Cloud is disabled by default. However, if you had enabled it and no longer need to use it, then you can disable DAS by deselecting the Enable DAS option from the Advanced Settings page.

About this task



Note: Disabling DAS from the **Advanced Settings** page does not affect the existing Database Catalogs and Virtual Warehouses. To remove and disable DAS from the Hive Virtual Warehouses, you must delete and recreate the Database Catalog and Virtual Warehouse.

Procedure

1. Log in to the Data Warehouse service as an Administrator.
2. Click Advanced Configuration to go to the **Advanced Settings** page.
3. Deselect the Enable DAS option.
4. Click Update.

What to do next

Delete and recreate a new Database Catalog and a Hive Virtual Warehouse.

Predefined Kerberos principals in Cloudera Data Warehouse Private Cloud

By default, Cloudera Data Warehouse (CDW) creates Kerberos principal names for Database Catalogs and Environments using the service hostname and the deterministic namespace name based on the name of the Database Catalog or Environment when you create a Database Catalog or an Environment. However, you can generate and provide the keytabs, if needed.

The service principals for CDW need to be the same as on the base cluster. For more information, see Customizing Kerberos principals in the CDP Private Base documentation.

By default, the host principals are generated programmatically. You can generate and provide the keytabs, but the hostnames in the Kerberos principals are fixed. CDW uses a deterministic namespace and environment IDs for the Kerberos principals.

When you specify an Environment or Database Catalog name, CDW appends a prefix as shown in the following table, as well as the Kerberos principal name based on them:

CDW entity	User-specified name	Namespace IDs with CDW-assigned prefix	Hive Kerberos principal name
Environment	my-test-env	env-my-test-env-default	hive/dwx-env-my-test-env@REALM.EXAMPLE.COM
Database Catalog	my-test-catalog	warehouse-my-test-catalog	hive/metastore-service.warehouse-warehouse-my-test-catalog.svc.cluster.local@REALM.EXAMPLE.COM
Virtual Warehouse	my-impala-warehouse	impala-my-impala-warehouse	NA

Note: The length of the namespace ID after CDW applies a prefix to the Environment or Database Catalog name, including the hyphen (-), should not exceed 63 characters. You can specify an Environment name 45 characters long and Database Catalog 53 characters long.

About the Refresh option

After changing settings and configurations, you often need to recreate Environments, Database Catalogs, and Virtual Warehouses. The Refresh option enables you to apply changes without the need to recreate them. Learn about the supported use cases in which you can use the Refresh option.

The Refresh option is available in the more options (⋮) menu at the Environment, Database Catalog, and Virtual Warehouse levels.

You might need to refresh the affected Database Catalogs and Virtual Warehouses when you add or update CA certificates for the LDAP server or update database settings such as host, port, database name, username, and password from the **Administration** page on Management Console. A refresh is also needed when you synchronize fresh configurations from the base cluster for the following components: Ozone, Hadoop, Hive, Ranger, and Atlas.

Note: You must refresh Database Catalogs first followed by Virtual Warehouses.

Third-party object storage support for Cloudera Data Warehouse Private Cloud

Cloudera Data Warehouse (CDW) can access object storage such as AWS S3 if the CDP Private Cloud base cluster is configured to connect to the object store. You can query Hive and Impala tables stored on object stores using Hue.

Note: Third-party object storage support is in technical preview and not recommended for production deployments. Cloudera recommends that you try this feature in test and development environments.

By default, when you activate an environment in CDW, all the hadoop.fs.s3a configurations (fs.s3a.*) are copied from the core-site.xml file present on the base cluster to the hadoop-core-site.xml file of the Hive and Impala metastore pods, enabling CDW to establish a connection to S3. The following four are the key configurations that must be present in the base cluster core-site.xml file:

- fs.s3a.access.key
- fs.s3a.secret.key

- fs.s3a.endpoint
- fs.s3a.connection.ssl.enabled

Important:

Because CDW uses all the hadoop.fs.s3a configurations from the base cluster, it is important that you fine-tune and debug these configurations on the base cluster before creating the CDW environment.

If you have installed the Private Cloud Data Services, including CDW, before fine-tuning the hadoop.fs.s3a configurations on the base cluster, then you must upload the Amazon server certificates referenced in the fs.s3a endpoint configuration on the Management Console Administration CA Certificates tab. Select Miscellaneous as the certificate type from the CA Certificate Type drop-down menu.

The fs.s3a.* configurations are read-only. You can view the fs.s3a.* configurations from the CONFIGURATION tab on the Database Catalog and Virtual Warehouse details page by selecting the hadoop-core-site.xml option from the Configuration files drop-down menu.

The Third-party S3 providers in private cloud option is enabled by default. You can disable CDW's access to S3 by deselecting the Third-party S3 providers in private cloud option from Advanced Configuration Advanced Settings page.

Note: Disabling CDW's access to the third-party S3 providers from the **Advanced Settings** page does not affect the previously created Database Catalogs and Virtual Warehouses. To disable access, you must delete and recreate the Database Catalog and Virtual Warehouse.

Using Ozone storage with Cloudera Data Warehouse Private Cloud

The topics in this section describe how to use Apache Ozone storage with Cloudera Data Warehouse (CDW) Private Cloud.

Note: Ozone support is in technical preview in CDW 1.4.1. Cloudera recommends that you use Ozone with CDW in test and development environments. It is not recommended for production deployments.

Creating a database on Ozone for Cloudera Data Warehouse Private Cloud Virtual Warehouses

Learn how to create a database on Ozone storage that can be used by Cloudera Data Warehouse (CDW) Private Cloud Hive or Impala Virtual Warehouses.

About this task

By default, the Hive metastore for Database Catalogs on CDW Private Cloud points to HDFS, but you can configure the Database Catalog to point to Ozone storage instead by using the following steps. These steps change the default Hive metastore location to Ozone.

Note: Ozone support is in technical preview in CDW 1.4.1. Cloudera recommends that you use Ozone with CDW in test and development environments. It is not recommended for production deployments.

Before you begin

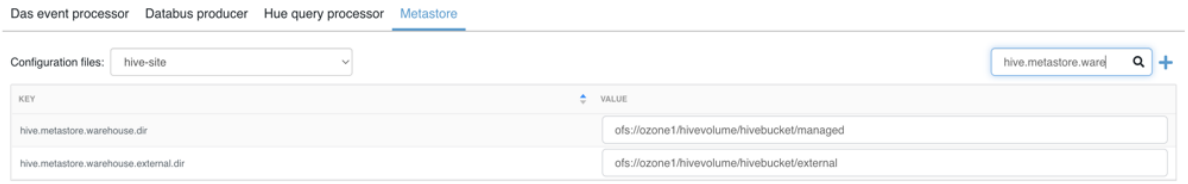
Before you re-configure the Database Catalog settings, make sure there are no running Virtual Warehouses associated with it. Either the Database Catalog has no associated Virtual Warehouses or you have suspended all the Virtual Warehouses associated with it.

Procedure

1. Use the following steps to change the Database Catalog setting:
   a) From the Management Console Private Cloud home page left menu, navigate to Data WarehouseOverview.
   b) In the Database Catalog tile, click ⋮ Edit .
   c) In the **Database Catalogs** detail page, click CONFIGURATIONS Metastore , and select hive-site from the Configuration files drop-down list.
   d) Search for the following configuration properties and update them to Ozone file system paths, which start with ofs:
      - hive.metastore.warehouse.dir
      - hive.metastore.warehouse.external.dir

      Note: For the Hive Table creation, the warehouse directory must be set at bucket level or directory level under the hive.metastore.warehouse.dir or hive.metastore.warehouse.external.dir parameters. For more information, see Changing the Hive warehouse location.

      Following is an example of these properties set for a Database Catalog:

      Note: The example values in the screenshot show the Hive warehouse locations in Ozone (set at a directory level) where Hive stores the tables. hivevolume represents the Ozone volume, hivebucket represents the Ozone bucket, and managed and external are directories where Hive stores the managed and external tables.
   e) Click Apply Changes and wait for the Database Catalog to finish applying changes.
2. Perform one of the following actions to get started working with a Virtual Warehouse:
   - Restart any associated Virtual Warehouses that you suspended before updating the Database Catalog properties by clicking the re-start icon in the upper right corner of the Virtual Warehouse tile on the Overview page.
   - Create a new Hive or Impala Virtual Warehouse associated with the updated Database Catalog.
3. Use Hue to create a database with your Virtual Warehouse. For details, see Querying data.

Results

After configuring the Database Catalog's Hive metastore to point to Ozone, you can create databases on Ozone with either an Impala or a Hive Virtual Warehouse.

Configuring Hive/Impala logging on Ozone for Cloudera Data Warehouse Private Cloud

This section describes how to configure Cloudera Data Warehouse (CDW) on Private Cloud to store Hive and Impala logs on Ozone storage.

You can configure CDW to store Hive and Impala logs on CDP Private Cloud storage components, such as Ozone. Ozone is a good choice to store these logs because:

- Ozone efficiently handles files regardless of their size.
- In addition to Ozone's built-in CLI interface, Ozone also supports the HDFS CLI and CLIs that are compatible with AWS clients.

- CDP Private Cloud uses fluentd to push application logs to the storage layer. Ozone is a supported logging "back-end" component and has a fluentd-compatible endpoint for collecting the logs.

Note: Ozone support is in technical preview in CDW 1.4.1. Cloudera recommends that you use Ozone with CDW in test and development environments. It is not recommended for production deployments.

Specify or create an Ozone bucket for Cloudera Data Warehouse Private Cloud logs

This topic describes how to specify an Ozone bucket to store Cloudera Data Warehouse (CDW) Private Cloud Hive and Impala logs.

About this task

You can either re-use the Ozone bucket that is automatically configured for storing Cloudera Machine Learning (CML) Private Cloud logs or create a new bucket to store CDW logs separately. The Ozone bucket used to store CML logs usually has a cdplogs- prefix.

Procedure

Use one of the following two methods depending on whether you want to use the existing CML log bucket or create a new one for CDW:

- To select an existing Ozone bucket, use the ozone sh bucket list command from the Ozone shell on your Private Cloud Base cluster. The following example shows how you can list buckets by the cdplogs- prefix:

```
ozone sh bucket list o3://ozone1/s3v --prefix=cdplogs
{
   "metadata" : { },
   "volumeName" : "s3v",
   "name" : "cdplogs-av-dwx-env-96c47aa9",
   "storageType" : "DISK",
   "versioning" : false,
   "creationTime" : "2020-08-01T18:29:08.686z",
   "modificationTime" : "2020-08-03T18:29:08.686z",
   "encryptionKeyName" : null,
   "sourceVolume" : null,
   "sourceBucket" : null
}
```

- To create a new bucket on Ozone, use the `ozone sh bucket create` command from the Ozone shell on your Private Cloud Base cluster. The following example shows how to create a new Ozone bucket named cdw-logs-bucket:

```
ozone sh bucket create o3://ozone1/s3v/cdw-logs-bucket
```

Important: Cloudera recommends that you use the hive user because this user automatically has create/read/write permissions on buckets that you create.

Update Cloudera Data Warehouse Private Cloud log configuration to point to Ozone

This topic describes how to configure Cloudera Data Warehouse (CDW) Private Cloud to store logs on Ozone.

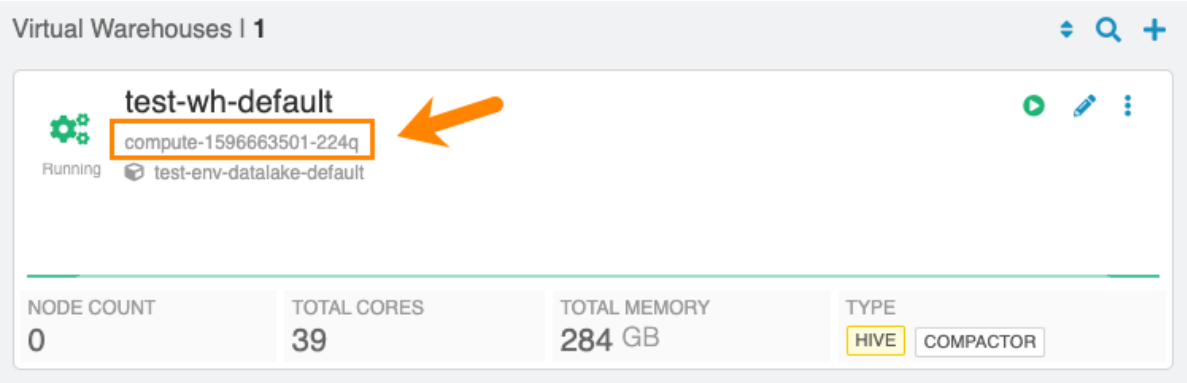
About this task

To configure CDW Private Cloud and the underlying OpenShift cluster to store Hive and Impala logs on Ozone, you must gather some information and prepare a block of code that you will insert into the Virtual Warehouse ConfigMap on the OpenShift pod. These preliminary steps are described in the following section.

Before you begin

Get the following information and prepare the block of code for the Virtual Warehouse ConfigMap before you start the steps of updating the configuration:

- Get the CDW namespace for your Virtual Warehouse:
  1. From the Management Console home page left menu, click Data Warehouse in the left menu. You are taken to the Overview page of CDW Private Cloud service.
  2. Locate the Virtual Warehouse you want to configure log storage for in the right-most column of the page, and locate the CDW namespace, which starts with compute- as shown below:



- Prepare the code block that must be pasted into the OpenShift ConfigMap:

  Here is an example:

```
<match **>
       @type s3
       @log-level debug
       aws_key_id <ACCESS-ID>
       aws_sec-key <SEC-KEY>
       s3_bucket <BUCKET-NAME>
       s3_endpoint <OZONE-S3-GATEWAY-ENDPOINT>
       ssl_verify_peer false
       s3_object_key_format
       "<WAREHOUSE_PREFIX>/warehouse/tablespace/external/hive/sys.db/logs
/dt=%Y-%m-%d/${path_tag}/%{time_slice}_${unique_file_key}.log.%{file_ext
ension}"
       time_slice_format %Y-%m-%d-%H-%M
       store_as gzip
       auto_create_bucket false
       check_apikey_on_start false
       force_path_style true
       check_bucket false
       check_object false
       <buffer path_tag, unique_file_key, time, warehouse>
        @type file
        path /tmp/fluentd-buffers/%{unique_file_key}-s3.buffer
        timekey 900 # minute precision for time_slice_format to have minu
te in file name
        timekey_use_utc true
        chunk_limit_size 265m
        flush_mode interval
        flush_interval "900s"
        flush_thread_count 8
        flush_at_shutdown true
       </buffer>
       <format>
        @type single_value
        message_key log
```

```
        add_newline true
      </format>
    </match>
```

In the above code block example:

- <*BUCKET-NAME*> indicates the name of the Ozone bucket used for storing the CDW Private Cloud logs.
- <*OZONE-S3-GATEWAY-ENDPOINT*> indicates the endpoint of the Ozone S3 Gateway. Get this value from the Ozone S3 Gateway Web UI page of Cloudera Manager.
- <*ACCESS_ID*> and <*SEC_KEY*> are the AWS access credentials for the Ozone S3 Gateway. Get these values by using the kinit -kt and the ozone s3 getsecre commands on the Private Cloud Base OpenShift cluster.

Procedure

1. Using OpenShift commands, view the OpenShift project for the pod where the CDW Private Cloud instance is running by specifying the CDW namespace for the Virtual Warehouse that you noted in the Before you begin section above.

   For example, if the CDW namespace is compute-1596663501-224q, you can view the OpenShift project with the following command:

   ```
   oc project compute-1596663501-224q
   ```

2. Open the ConfigMap for the Virtual Warehouse that is associated with the CDW namespace. For example:

   ```
   oc edit configmap warehouse-fluentd-config
   ```

   This command opens the ConfigMap in a separate editor that is similar to vi.
3. Replace the match section of the ConfigMap with the code block you prepared in the Before you begin section above, and then save your changes
4. Verify that the new configuration is correctly updated by running the following command:

   ```
   oc get namespace -o yaml | grep fluentd-status
   ```

   If the configuration is successfully updated, the value of the fluentd-status returns an empty string as shown in the following example:

   ```
   com.cloudera/fluentd-status: ""
         com.cloudera/fluentd-status: ""
         com.cloudera/fluentd-status: ""
         com.cloudera/fluentd-status: ""
   ```

Monitor Cloudera Data Warehouse Private Cloud logs on Ozone storage

This topic describes how to monitor Cloudera Data Warehouse (CDW) Private Cloud logs that are stored on Ozone.

About this task

You can use either the Ozone S3 Gateway Web UI in Cloudera Manager or run commands in a terminal window to monitor CDW logs.



Note: Because fluentd buffers the logs and then pushes them to the configured endpoint, Ozone might take up to 15 minutes to display the CDW logs.

Procedure

Use one of the following methods to monitor CDW logs in Ozone:

- Ozone S3 Gateway Web UI in Cloudera Manager:

  Navigate to the following URL:

  https://<*S3-GATEWAY-ENDPOINT*>/<*BUCKET-NAME*>?browser=true

  Where:

  - <*S3-GATEWAY-ENDPOINT*> indicates the endpoint of the Ozone S3 Gateway, which you can get from the Ozone S3 Gateway Web UI
  - <*BUCKET-NAME*> indicates the Ozone bucket where you are storing the CDW logs.
- Run the following command from the Ozone shell: ozone sh key list o3://<*OZONE.SERVICE.ID*>/s3v/<*BUCKET-NAME*>/ --prefix=<*WAREHOUSE-PREFIX*>

  Where:

  - <*OZONE.SERVICE.ID*> indicates the identifier used for your implementation of Ozone.
  - <*BUCKET-NAME*> indicates the name of the Ozone bucket where the CDW logs are stored.
  - <*WAREHOUSE-PREFIX*> indicates the Virtual Warehouse identifier.

Analyze Cloudera Data Warehouse Private Cloud logs stored on Ozone

This topic describes how to use Hue or Data Analytics Studio (DAS) to analyze Cloudera Data Warehouse (CDW) Private Cloud logs that are stored on Ozone.

About this task

You can use Hue to analyze Impala logs or DAS to analyze Hive logs.

Note:

You must use the Hue or DAS instance that corresponds to the Virtual Warehouse whose logs are saved on Ozone. To ensure that you use the correct instance, access Hue or DAS by using the drop-down menu in the upper right corner of the Virtual Warehouse tile.

Procedure

1. Using Hue or DAS, create an external table that points to the log data on Ozone:

   CREATE EXTERNAL TABLE <*TABLE-NAME*> LIKE sys.logs LOCATION 'o3fs://<*BUCKET-NAME*>.s3v.<*OZONE.SERVICE.ID*>/<*WAREHOUSE-PREFIX*>/warehouse/tablespace/external/hive/sys.db/logs';
2. Run the MSCK REPAIR TABLE command on the table you created in Step 1:

   MSCK REPAIR TABLE <*TABLE-NAME*>;

Results

After completing the above steps, you can use SQL queries to analyze the log data.

Changing delegation username and password

You specify the delegation username and password while activating an environment. You can change the delegation username or password from the Environment Details page.

Procedure

1. Log in to the Data Warehouse service as a DWAdmin.
2. Go to Environments ⋮ Edit CONFIGURATIONS .

3. Enter a new Delegation Username and/or Delegation Password.



Note: CDW supports only the following characters for specifying the Delegation Password:

- lowercase and uppercase alphabets
- numbers
- whitespace
- ! “ # $ % & ‘ ( ) * + , - . / : ; < = > ? @ [ ] ^ _ ` { | } ~

4. Click Apply Changes.

SSL-enabled endpoints for Virtual Warehouse clients in Cloudera Data Warehouse Private Cloud

In Cloudera Data Warehouse (CDW) Private Cloud 1.1, all client endpoints have been SSL-enabled. This requires that you configure the SSL certificates for client endpoints.

In CDW Private Cloud 1.1 and higher, client endpoints for web applications and Virtual Warehouse client URLs are SSL-enabled. The following endpoints use the OpenShift/Embedded Container Service cluster default certificate:

- Hue
- Data Analytics Studio (DAS) webapp
- Impala coordinator
- HiveServer2

Domain name changes

To use the OpenShift/Embedded Container Service cluster wildcard certificate, the DNS names have been changed. The environment ID sub domain from the domain name has been removed. This creates a flat DNA structure so the cluster wildcard certificate can be applied to the endpoints.

Generating a truststore for a self-signed certificate

You can query the service certificate and convert it to a JKS truststore using the following steps:

1. Retrieve the certificate:

```
$ openssl s_client -showcerts -connect hs2-my-cwh1.apps.cdw.mycloud.myfi
rm.com:443 -servername
hs2-my-cwh1.apps.cdw.mycloud.myfirm.com </dev/null|openssl x509 -outform
PEM > <MYCERTFILE>.pem
```

2. Convert the PEM file to a truststore. You will be prompted for a password.

```
$ keytool -import -alias hs2-my-cw1.apps.cdw.mycloud.myfirm.com -file
 <MYCERTFILE>.pem -keystore <MYCERT>.jks
```

Opening SSL-enabled connections with Database Catalog clients

The CDW Virtual Warehouse clients like beeline and impala-shell can open SSL-enabled connections as described in this section.

Beeline

A beeline connection can be created using a JDBC connection string. Specifying the username and password with the '-n' and the '-p' options returns an error. The beeline CLI prompts for credentials:

```
$ beeline
beeline> !connect
```

```
jdbc:hive2://hs2-my-cwh1.apps.cdw.mycloud.myfirm.com:443/default;transportMo
de=http;httpPath=cliservice;
     ssl=true;retries=3;sslTrustStore=<JKS-PATH>;trustStorePassword
=<***PASSWORD***>
Enter username for jdbc:hive2://hs2-my-cwh1.apps.cdw.mycloud.myfirm.com:443/
default:<MY-USER-NAME
Enter password for jdbc:hive2://hs2-my-cwh1.apps.cdw.mycloud.myfirm.com:443/
default:<********>
```

Important: The value for <*JKS-PATH*> is generated in the above section "Generating a truststore for a self-signed certificate."

impala-shell

The impala-shell CLI opens a TLS/SSL-enabled connection when you use the `--ssl` option. If `--ca_cert` is not set, impala-shell enables TLS/SSL, but does not validate the server certificate. Set the `--ca_cert` CLI option to the local path name that points to the third-party CA certificate, or to a copy of the server certificate in the case you have a self-signed server certificate:

```
$ impala-shell --protocol='hs2-http' -i "coordinator-my-iwh2.apps.cdw.myclou
d.myfirm.com:443" --ssl
```

OpenShift routes

OpenShift routes are used to expose the user-facing services in the CDW Private Cloud deployment. Route objects can perform edge TLS termination using the cluster-deployed certificate for the endpoints. If the cluster certificate must be rotated, the routes can pick up the new certificate automatically. It is not necessary to re-deploy or to manually configure the service in order to pick up the changes.

Debugging Impala Virtual Warehouses using Web UIs

You can use the Catalog Web UI, Coordinator Web UI, and the StateStore Web UI to debug Impala Virtual Warehouses in Cloudera Data Warehouse (CDW).

About this task

The Impala daemons (impalad, statestored, and catalogd) debug Web UIs, which can be used in CDP Runtime by using Cloudera Manager, is also available in the CDW service. In CDW service, the following Web UIs are provided:

- Impala Catalog Web UI

  This UI provides the same type of information as the Catalog Server Web UI in Cloudera Manager. It includes information about the objects managed by the Impala Virtual Warehouse. For more information about this debug Web UI, see .
- Impala Coordinator Web UI

  This UI provides the same type of information as the Impala Daemon Web UI in Cloudera Manager. It includes information about configuration settings, running and completed queries, and associated performance and resource usage for queries. For information about this debug Web UI, see .
- Impala StateStore Web UI

  This UI provides the same type of information as the StateStore Web UI in Cloudera Manager. It includes information about memory usage, configuration settings, and ongoing health checks that are performed by the Impala statestored daemon. For information about this debug Web UI, see

- Impala Autoscaler Web UI

  This UI gives you insight into Autoscaler operations, accessing log messages, and resetting the log level. The autoscaler Web UI includes information about the queries queued and running, executor groups, suspended calls, scale up/down calls, the autoscaler config, and the autoscaler logs.

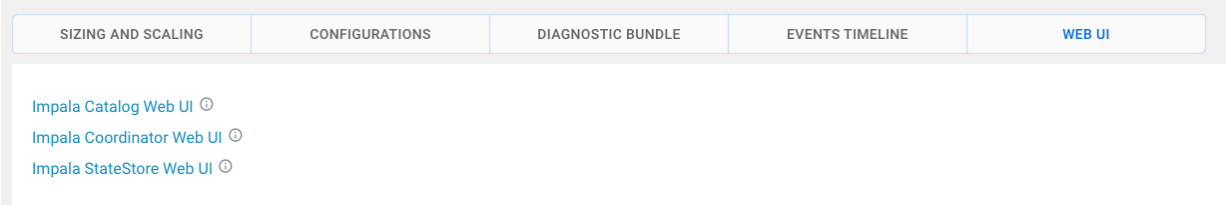
Required role: EnvironmentAdmin

Before you begin

Make sure that you note your CDP workload user name and have set a password for it in the User Management module of Management Console. You need to use your workload user name and its associated password to log into the debug Web UIs.

Procedure

1. In the CDW UI on the Overview page, locate the Impala Virtual Warehouse for which you want to view the debug UIs, and select Edit from the options menu on the tile. This launches the details page for this Virtual Warehouse.
2. In the **Virtual Warehouse** details page, select the WEB UI tab on the right. The list of debug Web UI links are displayed as shown in the following image:



3. Click a Web UI link corresponding to an Impala daemon that you want to debug.

   You are prompted to enter your workload user name and password.

Results

After you are authenticated, you can view the debug Web UI and use the information to help you troubleshoot issues with your Impala Virtual Warehouse.

Generating and downloading diagnostic bundles

Cloudera Data Warehouse (CDW) collects diagnostic data on workload logs, such as Impala Coordinator, Statefulset, CatalogD logs and stores it in the tmp directory on HDFS. You can download the logs using the Hue File Browser from the base cluster.

About this task

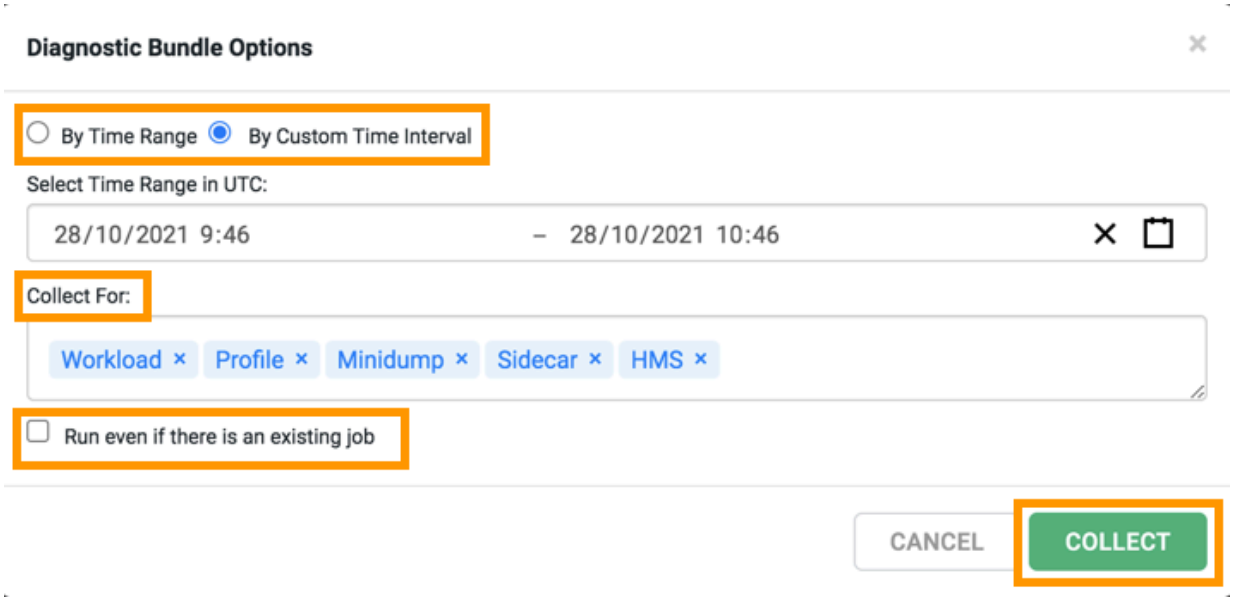
During the lifetime of a cluster, logs are continously written to the following directory on HDFS: [**WAREHOUSE-DIR**]/warehouse/tablespace/external/hive/sys.db/. When you click Collect Diagnostic Bundle from the CDW web interface, CDW collects the logs for the specified time interval and for the services that you select. These logs are compressed in a ZIP file format and stored in the tmp directory.

Attention: In 1.3.2 release of CDW Private Cloud, you can generate and download diagnostic bundles only for Impala.

Procedure

1. Log in to the CDW service as a DWAdmin.
2. Click the options drop-down menu on the Virtual Warehouse for which you want to collect the logs and click Collect Diagnostic Bundle.

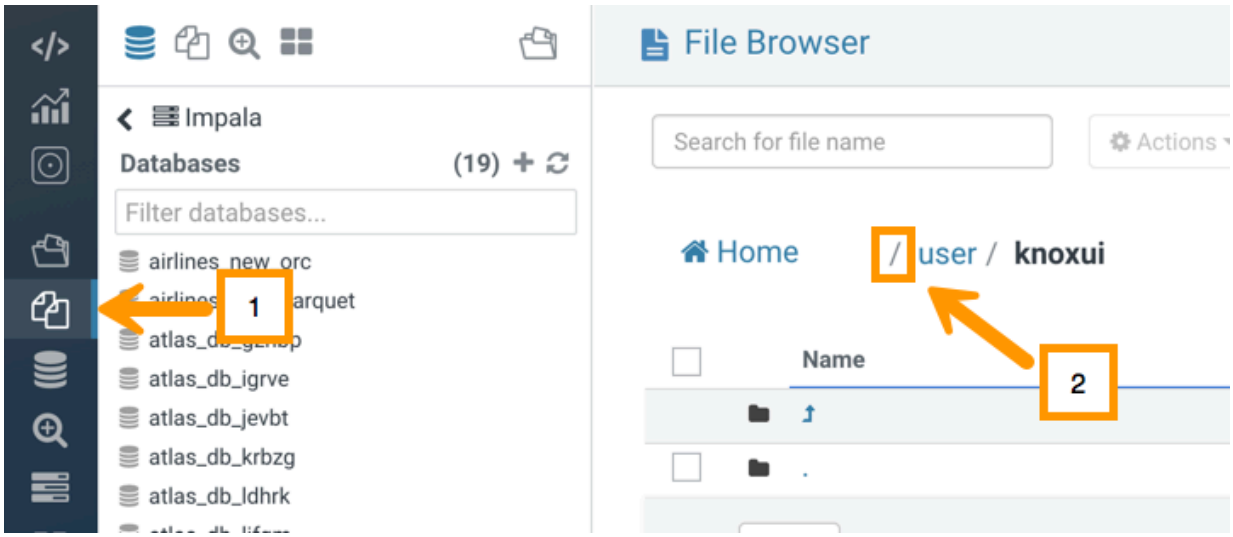
3. On the **Diagnostic Bundle Options** dialog box, select the time interval and the type of logs you want to collect and click COLLECT.



4. To view the status of the job and to obtain the HDFS location where the logs are stored, select Edit from the Virtual Warehouse options menu and go to the DIAGNOSTIC BUNDLE tab.

   The logs are collected and bundled under the /tmp/[***VIRTUAL-WAREHOUSE-ID-TIMESTAMP***].zip directory.

5. To access and download the logs, open the Hue service from the base cluster.
6. Go to the Hue File Browser and click the forward slash (/) before the user directory as shown in the following image:



The tmp directory is displayed. You can access and download the logs to your computer by clicking Download.

Configuring Impala Virtual Warehouses to encrypt spilled data in Cloudera Data Warehouse Private Cloud

If you have encrypted HDFS on the base CDP cluster, then Cloudera recommends that you configure an Impala Virtual Warehouse to write temporary data to disk during query processing in an encrypted format using the AES-256-CFB encryption for complete security.

About this task

In CDP Private Cloud, the temporary data is spilled to the local storage, the location of which is hard coded by the system.



Important: Impala does not selectively encrypt data based on whether the source data is already encrypted in HDFS. This results in at most 15 percent performance degradation when data is spilled.

Procedure

1. Log in to the Cloudera Data Warehouse service as an administrator.
2. Go to Impala Virtual Warehouse ⋮ Edit CONFIGURATIONS Impala coordinator and select flagfile from the Configuration files drop-down list.
3. Set the value of the disk_spill_encryption property to true.
4. Click APPLY.
5. Go to the Impala executor tab and select flagfile from the Configuration files drop-down list.
6. Set the value of the disk_spill_encryption property to true.
7. Click APPLY.
8. Restart the Impala Virtual Warehouse.

Creating custom pod configurations for Impala Virtual Warehouses

You can configure the resources used by Impala Virtual Warehouses in Cloudera Data Warehouse (CDW) Private Cloud environments to optimize Impala performance or to control resource usage in the environment.

About this task

When you create a Virtual Warehouse, CDW allocates standard resources to the Warehouses that are suitable for most workloads. You can control the size of the Virtual Warehouse at the time of creation by choosing the number of nodes to be used. By using custom pod configurations, you can also change the resources used by the critical Impala components, such as the coordinators, executors, and catalog daemons to pack a particular number of pods into a Kubernetes node or to create extra-large daemons to handle specific workloads.

Procedure

1. Log in to the Data Warehouse service as a DWAdmin.
2. Go to your environment and click ⋮ Edit .

   The **Environment Details** page is displayed.
3. Click the EDIT POD CONFIGURATIONS tab.

   A pod configuration is a named resource that is configured at the environment level.

4. Select one of the following two pod configuration options from the Select Pod Configuration section:
   - The Cdw Defaults option is selected by default. CDW uses default values for the pods if a specific pod configuration is not used.
   - Select the 1 x Node option for the allocation of most node resources found in the environment, to the Impala executors and coordinators.

   Cdw Defaults and 1 x Node are read-only options.
5. Click Copy Config to create and edit a new configuration with the option that you selected earlier as the basis.
   a) Specify the name for your configuration in the Cloned Config Name field.
   b) Enter a description for the new configuration in the Description for new config field.
   c) Click Create New Config.

      A new pod configuration is created, which you can now customize.
6. Specify the values for the following parameters under the Coordinator section:
   - Memory
   - Cpus
   - Xmx (maximum memory allocation pool for a Java Virtual Machine)
   - Xms (initial memory allocation pool for a Java Virtual Machine)
   - AC Slots (admission_control_slots flag)
   - Cache size (size of the data cache)
   - Scratch size (limit of Impala scratch space)
   - Overhead size (size for resources used by tools run by the containers)

   Note: By default, the total space allocated for scratch, cache, and overhead is 600GiB, with scratch=280GiB, cache=280GiB, and overhead=40GiB. You can allocate space for these components in the following format: [***QUANTITY***][***UNIT***].

   The quantity can be a decimal number, separated by a period (.). For example, 240.5GiB. The unit must be in GiB (gigibytes).

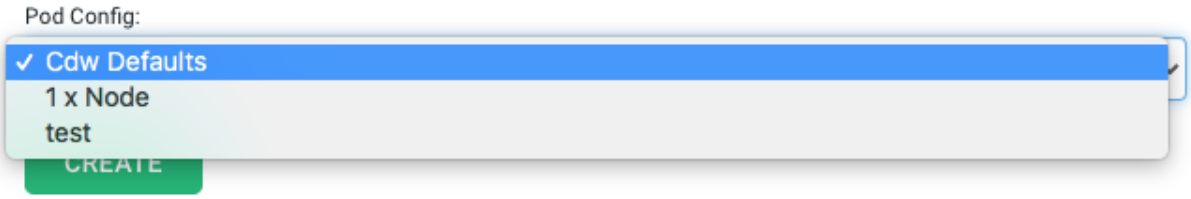
   1GiB (gibibyte) equals 1.074GB (gigabytes).
7. Specify the values for the following parameters under the Executor section:
   - Memory
   - Cpus
   - Xmx (maximum memory allocation pool for a Java Virtual Machine)
   - Xms (initial memory allocation pool for a Java Virtual Machine)
   - AC Slots (admission_control_slots flag)
   - Cache (size of the data cache)
   - Cache size (size of the data cache)
   - Scratch size (limit of Impala scratch space)
   - Overhead size (size for resources used by tools run by the containers)

   Note: By default, the total space allocated for scratch, cache, and overhead is 600GiB, with scratch=280GiB, cache=280GiB, and overhead=40GiB. You can allocate space for these components in the following format: [***QUANTITY***][***UNIT***].

   The quantity can be a decimal number, separated by a period (.). For example, 240.5GiB. The unit must be in GiB (gigibytes).

   1GiB (gibibyte) equals 1.074GB (gigabytes).

8. Specify the values for the following parameters under the Catalog section:
   - Memory
   - Cpus
   - Xmx (maximum memory allocation pool for a Java Virtual Machine)
   - Xms (initial memory allocation pool for a Java Virtual Machine)
9. Specify the values for the following parameters under the Default Settings section:
   - MaxQueryMemLimit
   - MinQueryMemLimit
   - mt_dop
10. Click Apply under the **EDIT POD CONFIGURATION** tab to save the custom settings.
    The “Configuration update initiated” message is displayed.
11. Click the Set as default configuration toggle button to make this a default pod configuration.
    This makes a pod configuration the default configuration at the environment level.
12. Click APPLY at the top of the Environment Details page.
    The new pod configuration becomes available in the Pod Config drop-down menu as shown in the following image. You can select this Impala pod configuration while creating a new Impala Virtual Warehouse:



Results

While adding a new Impala Virtual Warehouse, you can select the Pod Configuration to be used for resource allocation. The default value is "Cdw Defaults", but you can select other configurations available in your environment that you created using these steps.