

Troubleshooting

Date published:

Date modified:

CLOUDERA

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

- Troubleshooting Cloudera Data Warehouse..... 4**
 - Diagnostic bundles for Cloudera Data Warehouse and Kubernetes.....4
 - Downloading Cloudera Data Warehouse and Kubernetes diagnostic bundles..... 5
 - AWS environment fails to activate.....6
 - Opening Hue from Cloudera Data Warehouse causes an error.....8
 - Cleaning up old data to improve performance in Cloudera Data Warehouse..... 8
- Troubleshooting Impala.....10**
 - Downloading Impala diagnostic bundles..... 10
 - Accessing Impala Workload Logs..... 11
 - Locations of Impala Log Files in S3..... 11
 - Locations of Impala Log Files in Azure.....13
 - Analyzing Impala Virtual Warehouse Issues with the Events Timeline..... 15
 - Troubleshooting Impala..... 16
 - Debugging Impala Virtual Warehouses..... 17

Troubleshooting Cloudera Data Warehouse

Learn about common issues in Cloudera Data Warehouse, their cause, and the suggested steps to resolve them.

Diagnostic bundles for Cloudera Data Warehouse and Kubernetes

A diagnostic bundle captures information for troubleshooting and determining the root cause of problems in Cloudera Data Warehouse. Diagnostic bundles are available for troubleshooting your Virtual Warehouse, Database Catalog, Data Visualization, and environment/cluster.

You can obtain a diagnostic bundle collection for the following levels: Env (environment) DBC (Database Catalog), and VWH (Virtual Warehouse):

- Environment/cluster including Kubernetes resources
- DBC: Database Catalog including Kubernetes resources
- VWH:
 - Hive Virtual Warehouse including Kubernetes resources
 - Impala Virtual Warehouse including Kubernetes resources

The following information, and more, is included about Kubernetes resources:

- Pod
- Deployments
- CustomResource
- PVCs
- Statefulsets

At the Database Catalog and Environment level similar resource information and logs about the components are included in the bundle.

If you use Amazon CloudWatch and enable CloudWatch logs when you activate an AWS environment, or edit environment details, the following logs are included under their respective log streams and bundled at the ENV level:

- kube-scheduler
- kube-controller-manager
- kube-apiserver
- authenticator

For more information about enabling CloudWatch logs, see [activate an environment](#) from Cloudera Data Warehouse or [edit environment details](#). You must [add required permissions](#) to your IAM policy.

The diagnostic bundle includes a history of events at the cluster level. Although Kubernetes only preserves information about events of the past hour, the diagnostic bundle backs up events for the last 30 min or 1, 12, 24 hours, or a custom time interval you can specify.

Diagnostic Bundle Options for dw-impala ✕

☒ By Time Range
 ☐ By Custom Time Interval

Select A Time Range:

Last 30 Mins

▼

☐ Run even if there is an existing job

Collect

Cancel

You can also specify a custom time interval for collecting information.



Downloading Cloudera Data Warehouse and Kubernetes diagnostic bundles

To troubleshoot issues with Cloudera Data Warehouse, download diagnostic bundles ZIP files stored in your public cloud account. This capability is available on AWS environments only.

Before you begin


- Before you can download log files, you must, of course, run workloads on your Hive or Impala Virtual Warehouse to generate the logs.
- Obtain the DWAdmin role.

Procedure

1. Log in to the Cloudera web interface and navigate to the Cloudera Data Warehouse service.
2. Collect a diagnostic bundle for your environment, Database Catalog, or Virtual Warehouse as follows:
 - Environment bundle: In Environments, find your environment, click , and then click Collect Diagnostic Bundle.
 - Database Catalog bundle: Click Overview Database Catalog Diagnostic Bundle , and click Collect Diagnostic Bundle.
 - Virtual Warehouse bundle: Click Overview in the left navigation panel, select your Virtual Warehouse, click , and click Collect Diagnostic Bundle.
3. Select information about events within selectable time ranges, or select By Custom Time Interval, and specify the interval you want.
4. The Collect For field is available only for AWS environments with Amazon CloudWatch enabled. It allows you to choose the type of logs to include in the diagnostic bundle.
5. Click Collect.

After some time, depending on your cluster size and log sizes, but typically after 10 seconds, a message indicating completion appears indicating the diagnostic bundle is generated:

Collection of Diagnostic Bundle ... initiated.

6. Go to the **Overview** page, locate the environment, Database Catalog, or Virtual Warehouse from which you want to collect diagnostic data, and click  Edit Diagnostics Diagnostic Bundle .
7. In Diagnostic Bundles, copy the location of the diagnostic bundle ZIP file you want to download.

8. In the Amazon S3 management console, navigate to the location of the diagnostic bundle ZIP file, download the ZIP file, decompress it, and look at the troubleshooting information in the files.

Related Information

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

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

[Diagnostic bundle content](#)

AWS environment fails to activate

Learn how to resolve AWS environment activation failure in Cloudera Data Warehouse on cloud

When activating an AWS environment in Cloudera Data Warehouse on cloud, the following error message might be returned if your cloud resources reside in the us-east-1 region:

```
TemplateURL must reference a valid S3 object to which you have access.
```

Cause of the issue

For this region, the endpoint URL is incorrect and cannot load the CloudFormation template to create the AWS CloudFormation stack in your AWS account. Cloudera Data Warehouse uses these stack resources for Database Catalogs and Virtual Warehouses in Cloudera Data Warehouse.

Steps to resolve the issue

Use the reduced permissions mode feature for AWS environments to manually create the CloudFormation stack for Cloudera Data Warehouse. This feature enables you to manually create the CloudFormation stack resources from a template with a reduced set of IAM permissions. When you no longer need the environment, you must manually delete the CloudFormation stack resources in your AWS account.

1. Remove one of the permissions in your IAM permissions policy for the AWS account that you used to register the environment you want to activate for Cloudera Data Warehouse. For example, remove the `s3:CreateBucket` permission from your IAM permissions policy:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor0",
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupIngress",
        "acm:DeleteCertificate",
        "iam:RemoveRoleFromInstanceProfile",
        "s3:CreateBucket",
        "iam:CreateRole",
        "iam:AttachRolePolicy",
        "iam:PutRolePolicy",
        "dynamodb:DeleteTable",
        "ec2:DescribePlacementGroups",
        "rds:CreateDBSubnetGroup",
        "iam:AddRoleToInstanceProfile",
        "iam:DetachRolePolicy",
        "ec2:CreatePlacementGroup",
        "acm:RequestCertificate",
        "ec2:RevokeSecurityGroupEgress",
```

Remove →

Removing one of the required permissions from your IAM policy causes Cloudera Data Warehouse to display the reduced permissions mode option in the system dialog box when you activate your environment in Cloudera Data Warehouse.

- Follow the steps in [Activating AWS environments in reduced permissions mode](#).



Important:

In Step 4 of the [Activating AWS environments in reduced permissions mode](#) procedure, edit the pre-populated CloudFormation template URL as follows.

Change the URL from:

```
https://<BUCKETNAME>.s3-region.amazonaws.com/cf-templates/<CDW-CLUSTERID>-cf-template.yml
```

To:

```
https://<BUCKETNAME>.s3.amazonaws.com/cf-templates/<CDW-CLUSTERID>-cf-template.yml
```

- After you activate the environment and create the AWS CloudFormation stack resources in your AWS account, make sure that you apply the required tags to the stack that are listed in [Required tags for CloudFormation stacks created with reduced permissions mode](#).
- Add the s3:CreateBucket IAM permission back to your IAM permissions policy to make sure you have adequate permissions so Cloudera Data Warehouse can create CloudFormation stack resources for you when you activate environments later.

Deactivating environments created with the reduced permissions mode

When you no longer need the environment, you must manually delete the CloudFormation stack resources in the AWS Console by following the steps in [Deactivating AWS environments created with reduced permissions mode](#).

Opening Hue from Cloudera Data Warehouse causes an error

You need to be aware of some naming limitations when you create an environment. Observe the character limits for the Virtual Warehouse domain name.

Problem: After creating a Virtual Warehouse, you get an HTTP 500 error when you open Hue.

Probable Cause: The fully qualified domain name of your Virtual Warehouse, which includes the Virtual Warehouse name plus the environment name, has likely exceeded the limit: 64 characters.

Solution: Recreate the Virtual Warehouse using a name having a length that when added to the length of the environment name conforms to the limit.

Cleaning up old data to improve performance in Cloudera Data Warehouse

Some tables in Hue retain data indefinitely, resulting in slower performance or application crashes. Hue does not automatically clean up data from these tables. You can configure Hue to retain the data for a specific number of days and then schedule a cron job to clean up these tables at regular intervals for improved performance.

About this task

Consider cleaning up old data from the backend Hue database if you face the following problems while using Hue:

- Upgrade times out
- Performance is slower than expected
- Long time to log in to Hue
- SQL query shows a large number of documents in tables
- Hue crashes while trying to access saved documents



Important: The clean-up steps only delete the unsaved documents. Saved data and information are not cleaned up.

Before you begin

Back up your database before starting the cleanup activity. Check the saved documents, such as queries for a few users, to prevent data loss.



Note: The optimal number of documents that can be stored in a table is less than or equal to 30,000. Consider this number while specifying the cleanup interval.

Procedure

1. SSH into an active Hue instance.
2. Run the following command as the root user.

```
kubectl exec -it hue-0 -n <namespace> - /bin/bash
```

This command provides interactive access to the Hue pod (hue-0) within the Kubernetes cluster.

3. Enter the Hue Database Shell.

```
/opt/hive/build/env/bin/hue dbshell
```

This provides access to the Hue backend database shell. You can also note the sizes of the tables you want to clean up as a reference by running the following queries:

```
select count(*) from desktop_document;
select count(*) from desktop_document2;
select count(*) from beeswax_session;
select count(*) from beeswax_savedquery;
select count(*) from beeswax_queryhistory;
```

After completing the queries, exit the Hue database shell to continue.

4. Navigate to the required directory.

```
cd /opt/hive/build/env/bin
```

This ensures you are in the correct directory before running additional commands.

5. Remove outdated documents.

```
./hue desktop_document_cleanup --keep-days <x>
```

Removes outdated documents according to the specified retention period (x days).

6. To confirm document cleanup, repeat step 3 to validate if the table size has reduced.

```
SELECT COUNT(*) FROM desktop_document;
SELECT COUNT(*) FROM desktop_document2;
SELECT COUNT(*) FROM beeswax_session;
SELECT COUNT(*) FROM beeswax_savedquery;
SELECT COUNT(*) FROM beeswax_queryhistory;
```

7. Restart the Virtual Warehouse.

Troubleshooting Impala

Learn about common issues in Cloudera Data Warehouse Impala, their cause, and the suggested steps to resolve them.

Downloading Impala diagnostic bundles

Learn how to download diagnostic bundles to use for troubleshooting an Impala Virtual Warehouse in Cloudera Data Warehouse on cloud

About this task

To troubleshoot issues with your Impala Virtual Warehouses, download diagnostic bundles of log files for the sidecar containers that support Impala components and for the components themselves. These diagnostic bundles are ZIP files stored in your public cloud account.


Required role: DWAdmin

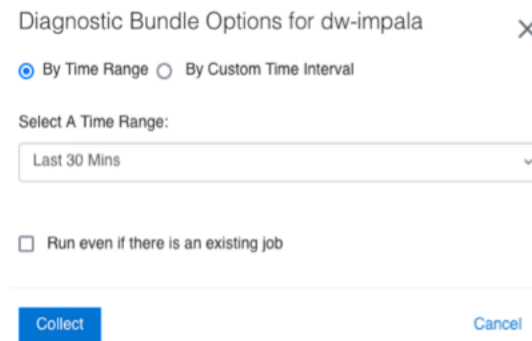
Before you begin

Before you can download log files, you must, of course, run workloads on your Impala Virtual Warehouse to generate the logs.

Procedure

1. Log in to the Cloudera web interface and navigate to the Cloudera Data Warehouse service.
2. Go to the Virtual Warehouses, locate the Virtual Warehouse from which you want to collect diagnostic data, and



click  Collect Diagnostic Bundle .



3. Set options that select which logs to generate for the diagnostic bundle.
 - By Time Range: Select a specific time range of log files to generate from the drop-down list, or you can choose a custom interval in the next option.
 - By Custom Time Interval: Select the start and end time from the drop-down list to define the specific time interval for the log files in the diagnostic bundle.
4. Click Collect to generate the bundle.

After some time, depending on your cluster size and log sizes, but typically after 10 seconds, a message indicating completion appears indicating the diagnostic bundle is generated:

```
Collection of Diagnostic Bundle for impala-mpvt initiated. Please go to
details page for more information.
```

5. Go to the Virtual Warehouses tab, locate the Virtual Warehouse from which you want to collect diagnostic data, and click  Edit Diagnostics .
6. Click copy-to-clipboard  to copy the path to the diagnostic bundle on S3/ABFS.
7. Paste the path to the diagnostic bundle into a text document, and navigate to the diagnostic bundle in AWS or Azure to download the ZIP file.
When you expand the diagnostic bundle ZIP file that you downloaded, directories appear for log files and a diagnostic-data-generator.log file, which contains troubleshooting information.

Related Information

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

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

[Diagnostic bundle content](#)

Accessing Impala Workload Logs

Describes how to locate Impala logs in S3 or Azure to diagnose some of the commonly encountered issues in Impala.

Using Impala Logs

The Impala logs record information about:

- Any errors Impala encountered.
- How Impala is configured.
- Jobs Impala has completed.

However, you can use the logs record information to troubleshoot only if the relevant logs are downloaded and then uploaded to a location where you can access them. To download the logs from S3 or Azure you must first identify the locations.

Locations of Impala Log Files in S3

This topic describes how to identify the Amazon S3 locations of Impala logs for the different Impala components.

About this task


The Cloudera Data Warehouse service collects logs from Impala Virtual Warehouses and uploads them to an Amazon S3 location. This S3 log location is configured under an external warehouse directory so that the logs are preserved even if the Virtual Warehouse they are collected from is destroyed.

Before you begin




To identify the location of the logs in S3, you must have the environment name, impala ID identifiers, and S3 bucket name.

Procedure

Finding the environment name, and impala ID identifiers



1. In the Data Warehouse service, expand the Environment column by clicking  .

- From the Overview page, note down the environment name, and impala ID identifiers.

Environments (103) Database Catalogs (10) Virtual Warehouses (28)					
<input type="text" value="Search by id or name"/> Status ▾ Type ▾ Environment ▾ Database Catalog ▾ ✕ Clear All					
Status	Name	Type	Version	CPU	Executor
 Good Health	dw-impala impala-1731012685-m2cw Impala ID dw-team-env-di-default dw-team-env Environment Name	 Impala	2024.0.19.0-266	62	

Identifying the bucket name

- In the Management console navigate to the Environment page.
- On the Environment page, click on your Environment and then on the Summary tab.
- On the Logs Storage and Audits section. You can find your log bucket and prefix under the Storage Location. This name is required to identify the S3 location of the logs.

 Logs Storage and Audits	
Storage Location:	s3a://dw-bucket/logs
Instance Profile:	arn:aws:iam::999999999999:instance-profile/admin-role
 Backup Storage	
Storage Location:	s3a://dw-bucket/logs
Instance Profile:	arn:aws:iam::999999999999:instance-profile/admin-role

Log locations in S3

- Now that you have identified the S3 bucket name, environment name, and impala ID identifiers, use the following prefix to find the logs generated by specific components in the following directories. Use the different directories listed here to view Impala/Hue logs.

```
PREFIX =
s3://<s3_bucket_name>/<prefix>/cdw/clusters/<environment_name>/logs/dt=
<date_stamp>/ns=<impala_ID>/
```

Impala component	S3 directory location
impalad	PREFIX + "app=impala-executor-log"
catalogd	PREFIX + "app=catalogd-log"
coordinator	PREFIX + "app=coordinator-log"
auto-scaler	PREFIX + "app=impala-autoscaler-log"
Hue	PREFIX + "app=huebackend-log" PREFIX + "app=hue-huedb-create-job-log" PREFIX + "app=huefrontend-log"

statestored	PREFIX + "app=statestored-log"
-------------	--------------------------------

The impalad executor logs for 25 August 2024 are located in the following S3 location:

```
PREFIX = s3://<s3_bucket_name>/<prefix>/cdw/clusters/<environment_name>/
logs/dt=2024-08-25/app=impala-executor-log/
```

In the above location, you can find multiple logs that were generated on the specified day.

Impala Minidumps

7. Impala minidumps can be found under the debug-artifacts/impala directory

```
/cdw/clusters/<environment_name>/logs/debug-artifacts/impala/<impala_ID>/
minidump/<pod_name>/
```

Impala Query Profiles

8. Impala query profiles are written in thrift encoded format in this location:

Impala component	S3 directory location
Impala query profiles	PREFIX + "app=impala-profiles"

Use the binary tool to decode thrift to text. This binary tool is provided with the upstream runtime Impala 4.0 as a docker image. Run the following command to use this tool.

```
docker run -i apache/impala:4.0.0-impala_profile_tool < name of the thrift
encoded file to decode
```

You can use the docker image available [here](#) to use this decoding tool.

Locations of Impala Log Files in Azure

This topic describes how to identify the Azure locations of Impala logs for the different Impala components.

About this task

The Cloudera Data Warehouse service collects logs from Impala Virtual Warehouses and uploads them to the Azure storage account that was provided while registering the Environment. This ABFS log location is configured under an external warehouse directory so that the logs are preserved even if the Virtual Warehouse they are collected from is destroyed.

Before you begin

To identify the location of the logs in Azure, you must have the environment name and impala ID identifiers and to access the logs from the Azure Portal you must know your storage account name.

Procedure

Finding the environment name, impala ID identifiers

1. In the Cloudera Data Warehouse service, go to the Virtual Warehouse tab.

- From the Overview page, note down the environment name and impala ID identifiers.

Environments (103) Database Catalogs (10) <u>Virtual Warehouses (28)</u>					
<div> <input type="text" value="Search by id or name"/> <div> <div>Status</div> <div>Type</div> <div>Environment</div> <div>Database Catalog</div> <div>×</div> <div>Clear All</div> </div> </div>					
Status	Name	Type	Version	CPU	Executor
<div> <div>Good Health</div> </div>	<div>dw-impala</div> <div> <div>Impala ID</div> <div>Impala-1731012685-m2cw</div> <div>dw-team-env</div> <div>Environment Name</div> </div>	<div>Impala</div>	2024.0.19.0-266	62	

Retrieving your storage account name

- In the Management Console navigate to the Environments page.
- On the Environments page, click on your Environment and click on the Summary tab.
- Scroll down to the Logs Storage and Audits section.

<div> <div>Logs Storage and Audits</div> <div>Storage Location: abfs://log@cdwstorage.dfs.core.windows.net</div> </div>	
<div> <div>Backup Storage</div> <div>Storage Location: abfs://log@cdwstorage.dfs.core.windows.net</div> </div>	

Note down your storage account name.

Accessing the different directories in the Azure Portal

- Log in to the Azure Portal and search for your storage account name using the Search bar.
- On the Overview page of your storage account, click on the Containers menu.
- Click on the file system you used during the Environment registration.



Note: You need to enable the firewall rules, click on the Firewalls and virtual networks menu, and set Allow access to “All networks”, then save the changes to access the file system.

Microsoft Azure	
<div> <div>Home > Storage accounts > cdwstorage Containers</div> <div> <div>logs</div> <div>Container</div> </div> <div> <div>Search</div> <div>Upload</div> <div>Add Directory</div> <div>Refresh</div> <div>Rename</div> <div>Delete</div> <div>Change tier</div> <div>Acquire lease</div> <div>Break lease</div> <div>Give feedback</div> </div> <div> <div>Overview</div> <div>Diagnose and solve problems</div> <div>Access Control (IAM)</div> <div>Settings</div> </div> <div> <div>Authentication method: Access key (Switch to Microsoft Entra user account)</div> <div>Location: logs</div> <div>Search blobs by prefix (case-sensitive)</div> <div> <div>Name</div> <div> <input type="checkbox"/> cdw <input type="checkbox"/> cluster-backups <input type="checkbox"/> cluster-logs <input type="checkbox"/> clusters </div> </div> </div> </div>	

Log locations in ABFS

- Use the environment name, and impala ID identifiers, in the following prefix to find the logs generated by specific components in the following directories. Use the different directories listed here to view Impala/Hue logs

PREFIX =

```
/cdw/clusters/<environment_name>/logs/dt=<date_stamp>/ns=<impala_ID>/
```

Impala component	ABFS directory location
impalad	PREFIX + "app=impala-executor-log"
catalogd	PREFIX + "app=catalogd-log"
coordinator	PREFIX + "app=coordinator-log"
auto-scaler	PREFIX + "app=impala-autoscaler-log"
Hue	PREFIX + "app=huebackend-log" PREFIX + "app=hue-huedb-create-job-log" PREFIX + "app=huefrontend-log"
statestored	PREFIX + "app=statestored-log"

The impalad executor logs for 25 August 2024 are located in the following ABFS location:

```
PREFIX = /cdw/clusters/<environment_name>/logs/dt=2024-08-25/app=impala-executor-log/
```

In the above location, you can find multiple logs that were generated on the specified day.

Impala Minidumps

10. Impala minidumps can be found under the debug-artifacts/impala directory

```
/cdw/clusters/<environment_name>/logs/debug-artifacts/impala/<impala_ID>/minidump/<pod_name>/
```

Impala Query Profiles

11. Impala query profiles are written in thrift encoded format in this location:

Impala component	ABFS directory location
Impala query profiles	PREFIX + "app=impala-profiles"

Use the binary tool to decode thrift to text. This binary tool is provided with the upstream runtime Impala 4.0 as a docker image. Run the following command to use this tool.

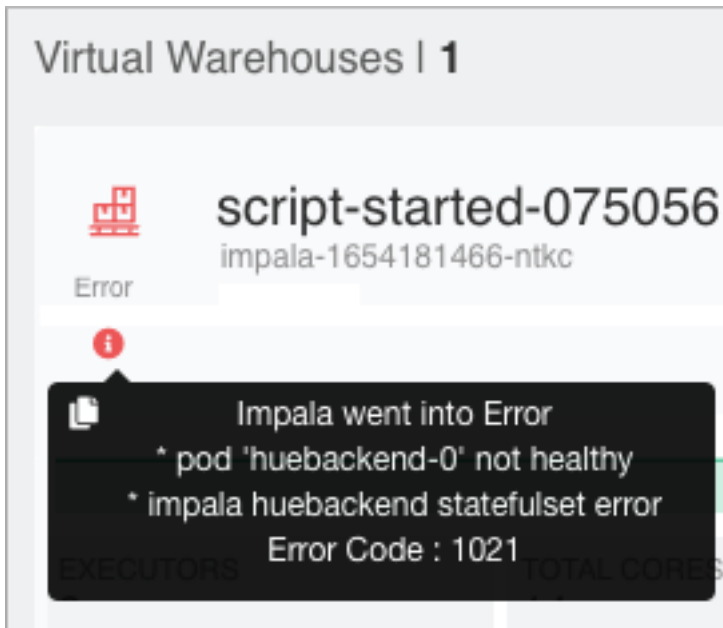
```
docker run -i apache/impala:4.0.0-impala_profile_tool < name of the thrift encoded file to decode
```


You can use the docker image available [here](#) to use this decoding tool.

Analyzing Impala Virtual Warehouse Issues with the Events Timeline

The Cloudera Data Warehouse server gathers information about problems that cause Virtual Warehouse errors and displays a tooltip in the Virtual Warehouse tile. Also, from the user interface, you can get details about events that can help you solve problems.

In the Virtual Warehouse tile, you now see tooltips. For example:



Go to the **Overview** page and navigate to Virtual Warehouses  Edit Diagnostics Events Timeline , you expand timeline items to get details that can help you troubleshoot problems.

Troubleshooting Impala

This topic describes the general troubleshooting procedures to diagnose some of the commonly encountered issues in Impala.

Symptom	Explanation	Recommendation
Impala takes a long time to start.	Impala instances with large numbers of tables, partitions, or data files take longer to start because the metadata for these objects is broadcast to all <code>impalad</code> nodes and cached.	Adjust timeout and synchronicity settings.
Joins fail to complete.	There may be insufficient memory. During a join, data from the second, third, and so on sets to be joined is loaded into memory. If Impala chooses an inefficient join order or join mechanism, the query could exceed the total memory available.	<p>Start by gathering statistics with the <code>COMPUTE STATS</code> statement for each table involved in the join.</p> <p>Consider specifying the <code>[SHUFFLE]</code> hint so that data from the joined tables is split up between nodes rather than broadcast to each node.</p> <p>If tuning at the SQL level is not sufficient, add more memory to your system or join smaller data sets.</p>
Queries return incorrect results.	Impala metadata may be outdated after changes are performed in Hive.	After inserting data, adding a partition, or other operation in Hive, refresh the metadata for the table with the <code>REFRESH</code> statement.
Attempts to complete Impala tasks such as executing <code>INSERT SELECT</code> statements fail. The Impala logs include notes that files could not be opened due to permission denied.	This can be the result of permissions issues. For example, you could use the Hive shell as the <code>hive</code> user to create a table. After creating this table, you could attempt to complete some action, such as an <code>INSERT SELECT</code> on the table. Because the table was created using one user and the <code>INSERT SELECT</code> is attempted by another, this action may fail due to permissions issues.	Ensure the Impala user has sufficient permissions to the table that the Hive user created.

Symptom	Explanation	Recommendation
Impala fails to start up, with the <code>impalad</code> logs referring to errors connecting to the statestore service and attempts to re-register.	A large number of databases, tables, partitions, and so on can require metadata synchronization, particularly on startup, that takes longer than the default timeout for the statestore service.	Configure the statestore timeout value and possibly other settings related to the frequency of statestore updates and metadata loading.

Debugging Impala Virtual Warehouses

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

Table level events

In addition to global metrics described below, the following table metrics are available for debugging an Impala Virtual Warehouse:

- `avg-events-process-duration`
- `events-consuming-delay-ms`

`avg-events-process-duration` metric

This metric represents the sum of the time for processing all events. This metric is helpful to identify the average duration of processed events on the table and to identify which tables are causing the event-processor to lag behind. As a temporary workaround, you can disable event processing on that table. You can set the metric collection period to 1 minute, 5 minutes, and 15 minutes duration:

- `avg-events-process-duration-1min-rate`
Exponentially weighted moving average (EWMA) of number of events processed in last 1 min
- `avg-events-process-duration-5min-rate`
Exponentially weighted moving average (EWMA) of number of events processed in last 5 min
- `avg-events-process-duration-15min-rate`
Exponentially weighted moving average (EWMA) of number of events processed in last 15 min

Metric output looks something like this:

  localhost:25020/table_metrics?name=default.tb1



```
events-process-duration:
  Count: 2
  Mean rate: 0.0109
  1 min. rate: 0.0066
  5 min. rate: 0.0048
  15 min. rate: 0.002
  Min (msec): 24
  Max (msec): 33
  Mean (msec): 28.9089
  Median (msec): 33.4825
  75th-% (msec): 33.4825
  95th-% (msec): 33.4825
  99th-% (msec): 33.4825

hms-load-tbl-schema:
  Count: 1
  Mean rate: 0.0055
  1 min. rate: 0.0108
  5 min. rate: 0.1116
  15 min. rate: 0.1647
```

events-consuming-delay-ms metric

This metric represents the time difference between creating an event in the metastore and processing an event. Using this metric, you can gauge how long the event processor is lagging.

Metric output looks something like this:

  localhost:25020/events

```
Mean (msec): 67.9801
Median (msec): 78.3973
75th-% (msec): 78.3973
95th-% (msec): 78.6436
99th-% (msec): 78.6436
```

events-consuming-delay:

```
Count: 12
Mean rate: 0.0342
1 min. rate: 0.0017
5 min. rate: 0.0146
15 min. rate: 0.0095
Min (msec): 2000
Max (msec): 10000
Mean (msec): 3472.5061
Median (msec): 2000
75th-% (msec): 4000
95th-% (msec): 8000
99th-% (msec): 10000
```

About this task

The Impala daemons (impalad, statestored, and catalogd) debug Web UIs, which can be used in by using , is also available in the service. In 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 . It includes information about the objects managed by the Impala Virtual Warehouse. For more information about this debug Web UI, see [Debug Web UI for Catalog Server](#).

- Impala Coordinator Web UI

This UI provides the same type of information as the Impala Daemon Web UI in . 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 [Debug Web UI for Impala Daemon](#).

- Impala StateStore Web UI

This UI provides the same type of information as the StateStore Web UI in . 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 [Debug Web UI for StateStore](#).

- Impala Autoscaler Web UI


This UI gives you insight into autoscaler operations (regular as well as workload-aware autoscaling), 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.

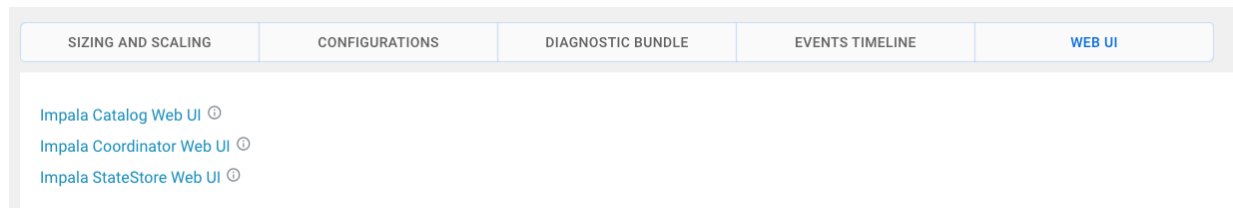
Required role: EnvironmentAdmin

Before you begin

Make sure that you note your workload user name and have set a password for it in the User Management module of . You need to use your workload user name and its associated password to log into the debug Web UIs. For more information, see [Setting the workload password](#) in the documentation set.

Procedure

1. In the UI on the Overview page, locate the Impala Virtual Warehouse for which you want to view the debug UIs, and select  Edit . The **Virtual Warehouse Details** page is displayed.
2. In the **Virtual Warehouse Details** page, select the Web UI tab. 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.