Cloudera Data Warehouse Private Cloud 1.5.2

Data Warehouse Release Notes

Date published: 2020-08-17
Date modified: 2023-11-02

CLOUDERA

Legal Notice

Contents

What's new in Cloudera Data Warehouse Private Cloud

Learn about the new features in Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services 1.5.2.

CDW simultaneously supports LDAP and Kerberos authentication for Hive Virtual Warehouses

Earlier, you could select either LDAP or Kerberos for authenticating users to Hive Virtual Warehouses while creating a new Virtual Warehouse. Starting with 1.5.2, CDW simultaneously supports both LDAP and Kerberos authentication mechanisms. See Authenticating users in CDW Private Cloud.

Group-level access control for Hive Virtual Warehouse (Preview)

CDW enables you to allow one or more user groups to access a particular Hive Virtual Warehouse, similar to Impala. As a result, only specific users can connect to a Virtual Warehouse, from all supported channels (Hue, Beeline, JDBC, or other Business Intelligence tools). The option to specify user groups when creating a new Virtual Warehouse is disabled by default. See Enabling warehouse-level access control for Hive and Impala in CDW Private Cloud.

Ability to use quota-managed resource pools

You can now assign quota-managed resource pools for environments, Data Catalogs, Virtual Warehouses, and Data Visualization instances in CDW. To learn more about quota management in CDP Private Cloud Data Services, see Managing cluster resources using Quota Management. To enable the use of quota management for CDW entities, see Enabling quota management in CDW Private Cloud.

About Iceberg support in CDW Private Cloud Data Services 1.5.2

Apache Iceberg is generally available (GA) for Impala and Hive LLAP in CDW Private Cloud when deployed on CDP Private Cloud Base version 7.1.9 or higher. This applies to CDW LLAP and Impala, and CDP Private Cloud Base with Spark 3, Impala, Nifi, and Flink. See Apache Iceberg in Cloudera Data Platform.

Apache Iceberg is in Technical Preview for Impala and Hive LLAP in CDW Private Cloud when deployed on CDP Private Cloud Base version 7.1.7 SP2 or 7.1.8, because interoperability requirements between CDP Private Cloud Base and CDW Private Cloud are not met. Tables that are converted to Iceberg Table format can only be accessed through CDW Impala and LLAP.

Note: Preview features are not recommended for production deployments. Cloudera recommends that you use Apache Iceberg integration with CDW in test or development environments.

Note: On CDP Private Cloud Base 7.1.9, Iceberg table entities are not created in Atlas. If you are using CDP Private Cloud Base 7.1.7 SP2 or 7.1.8, then you must manually upload the Iceberg model file 1130-iceberg_table_model.json in to the /opt/cloudera/parcels/CDH/lib/atlas/models/1000-Hadoop directory. See the known issue "CDPD-59413: Unable to view Iceberg table metadata in Atlas".

Added support for collecting CDW audit logs

Audit logging is now supported for CDW. In Private Cloud, audit logs from the Control Plane service and CDW are sent to the OpenTelemetry (OTEL) collector. The audit events do not persist. You can configure the OTEL collector to send data to external systems such IBM Guardian by using the syslog OTEL exporter. The OTEL collector is located in the Control Plane namespace, which is a pod with the following name format:

```
cdp-release-opentelemetry-collector-<UNIQUE-ID>
```

You can use the event data for troubleshooting issues. For more information, see Auditing Control Plane activity. For the list of the audit events collected for CDW, see Cloudera Data Warehouse audit events.

Unified timezone on the CDP Private Cloud Base and Data Services clusters

CDW logs, diagnostic bundles, and time-related SQL queries can now use the timezone specified in Cloudera Manager. On OCP, you must manually install the third-party webhook called K8TZ helm chart before the CDP Private Cloud Containerized Cluster setup step. For more information, see ECS unified time zone.

> **Note:** The timestamp in the filenames of the files inside a diagnostic bundle is as per the timezone configured in Cloudera Manager. However, if you directly list the files in HDFS, the timestamp in the filenames is as per UTC, but the file contents are as per the timezone configured in Cloudera Manager.

Added support for using dedicated worker nodes

On a CDP Private Cloud Data Services cluster, you can taint the nodes having the specialized hardware for specific workloads, such as CDW or Cloudera Machine Learning (CML). For example, CML requires higher GPU power and CDW requires more local storage (either SSD or NVME). If you have tainted the worker nodes in Cloudera Manager, then you can enable the Use dedicated nodes for executor option while activating an environment in CDW or by editing an existing environment. CDW and other data services can then schedule the Hive and Impala executor and coordinator pods on these dedicated worker nodes tainted exclusively for CDW. Other pods such as Data Visualization or Hive MetaStore (HMS) can only be scheduled on any of the undedicated nodes available within the cluster. For more information, see Scheduling executor pods on dedicated worker nodes in CDW Private Cloud.

Ability to view Impala query details and query profile in Hue

You can now view Impala query details, query plan, execution summary, and query metrics on the new **Impala Queries** tab on the **Job Browser** page using Hue in CDW, and use this information to tune and optimize your queries. You can also compare two queries and view all query details side-by-side. You can also view the Impala query profile and other related details on the new **Impala** tab on the **Job Browser** page. See Viewing Impala query details.

CDW is supported on a multiple-base-cluster deployment

You can configure one Embedded Container Service (ECS) cluster to work with multiple CDP Private Cloud Base clusters managed by separate instances of Cloudera Manager. CDP Private Cloud Data Services 1.5.2 onward, CDW is supported in this deployment configuration. For more information, see Configuring multiple Base clusters with one ECS cluster.

Improvements
Improved pod placement policy

> When scheduling HiveServer2 (HS2), executors, and coordinators in Impala and Hive Virtual Warehouses, CDW now considers rack configurations specified in Cloudera Manager. If sufficient resources are available, CDW prefers the same racks for scheduling HS2, executor, and coordinator pods. For more information, see Pod placement policy and rack awareness in CDW Private Cloud.

Improved third-party integration with CDW on Private Cloud

> Cloudera has improved the tagging and labeling of CDW entities so that the third-party applications can write Kubernetes injectors to add custom services to CDW. This allows the add-on services built on top of CDW to work in tandem without needing to modify the CDW Docker images. To integrate your services with CDW, these services must provide their own Helm charts and by using node labels, these services can run as sidecar containers to the CDW pods. For the list of available labels, see List of labels for third-party integration.

CLI client drivers are packaged with CDW

> Previously, the CLI client drivers such as JDBC and ODBC drivers for Hive and Impala, and Beeline CLI client were hosted in a utility bucket on Amazon S3, making it difficult to access and download in private cloud environments. These client drivers are now packaged and distributed with CDW. You can download them from the Resources and Downloads tile.

Added support for unsecured LDAP for Impala

Earlier, CDW required you to use secured LDAP (LDAPS) for authenticating users accessing an Impala Virtual Warehouse. You can now use an unsecured LDAP server for authentication. CDW now uses the LDAP configurations that you have configured in the CDP Management Console.

"fe_service_threads" configuration is not copied from the base cluster to CDW

In CDP Private Cloud Data Services 1.5.1, the "fe_service_threads" configuration, which is used to specify the maximum number of concurrent client connections or threads allowed to serve client requests in Impala, was copied from the CDP Private Cloud base cluster to CDW along with its set value (typically 64). This degraded the performance. Starting with the 1.5.2 release, this configuration is no longer copied from the base cluster. You can set the value of the "fe_service_threads" configuration based on your requirements (recommended value is 96 or higher). See Configuring "fe_service_threads" in CDW Private Cloud.

Support for non-TLS-enabled HMS database

Earlier, you had to manually run a script for establishing unsecured connections to the Hive MetaStore (HMS) database (non-TLS-enabled). This process has been automated in the CDP Private Cloud Data Services 1.5.2 release. CDW automatically configures a secured or unsecured connection (with or without TLS) to the HMS database according to the configurations present on the CDP Private Cloud Base cluster.

High Availability (HA)-compatible Hive delegation token store automatically set for Database Catalogs

Storage for the Kerberos delegation token is defined by the hive.cluster.delegation.token.store.class property. CDW automatically sets the value of this property to org.apache.hadoop.hive.thrift.DB TokenStore for the Database Catalog, which is always created in the High Availability mode.

Known issues and limitations in Cloudera Data Warehouse Private Cloud

This section lists known issues and limitations that you might run into while using the Cloudera Data Warehouse (CDW) service in CDP Private Cloud Data Services.

General known issues

This topic describes the general service-wide known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.2

DWX-16826: Executor pods do not get scheduled on the same node after the scale-down

When the executor pods scale down and scale back up, they make an attempt to get scheduled on the same node because their PersistentVolumeClaim (PVC) is bound to a Persistent Volume on that node. This can cause issues if other pods are scheduled on the node on which the executor pod is attempting to get scheduled between a scale-down and scale-up cycle, taking away the resources from the executor.

1. Cordon the node on which the executor pod is attempting to get scheduled, which evicts the other pods that got scheduled on this particular node.
2. Wait until the pods that got evicted are scheduled on other available nodes within the cluster.
3. Uncordon the node, so that the executor pod can get be scheduled again on the specific node.

DWX-17179: Hue backend and Impala catalog pods are scheduled on the same node in an HA deployment

You may notice that multiple replicas of Hue backend, frontend, and Impala catalog pods get scheduled on the same node in HA mode.

You can manually move pods to other nodes by adding anti-affinity rules to the deployments.

VIZ-2269: Issue with impersonation in CDW Impala connections

> You may see the following error while creating or editing the data connection to a CDW Impala or Hive Virtual Warehouse and using the connection details auto-populated from the CDW Warehouse drop-down: User <username> is not authorized to delegate to <username>. This happens because the Impersonation and Trusted Impersonation options are both enabled. This affects CDV 7.1.6.

> If the message appears when creating a new data connection, try refreshing the page. This will reset the Impersonation and Trusted Impersonation options. If the message appears on the **Edit Data Connection** modal, then copy the hostname, username, HTTP path, and other values present on the **Advanced** tab, and manually edit the existing CDW connections. A manual edit will typically not trigger this bug.

Known issues identified in 1.5.1

DWX-15142 Character restriction on environment name when using FreeIPA server version 4.9.8 and higher

> FreeIPA is supported as an authentication mechanism starting with the 1.5.1 release. If you are using FreeIPA version 4.9.8 and higher, then note that the host names are limited to 64 characters. Because the environment name is part of the host name, the environment name must not exceed 17 characters.

> None.

Known issues identified before 1.4.1

DWX-10403: Executor pods get stuck in pending state with a warning

> In rare circumstances, when Impala or Hive executors start up either due to autoscaling or by manually restarting the executors, the pods may get stuck in a pending state with a warning such as "volume node affinity conflict". This happens due to a race condition in the storage class that provides local volumes.

> Restart the pods so that they can be rescheduled on new nodes with enough resources.

DWX-8502: HMS health check does not check port 9083

> The HMS health check script does not check the health of its service port 9083 and may provide incorrect health status.

> None.

Upgrade-related known issues

This topic describes the upgrade-related known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

Known issues identified before 1.5.2

DWX-15571: tez.history.logging.proto-base-dir and hive.hook.proto.base-directory locations change after the runtime upgrade

> The values of following Hue Query Processor properties change after you upgrade the Database Catalog and Virtual Warehouses to the latest version:

> • tez.history.logging.proto-base-dir
> • hive.hook.proto.base-directory

> CDW changes these locations to match the base cluster configuration for hive.metastore.warehouse .external.dir. You may also see the FileNotFoundException: File does not exist error in the Hue Query Processor logs.

> **Note:** This known issue does not cause any change in the functionality.

To use the old location and to avoid seeing the "File does not exist" error in the logs, you must update the Hue Query Processor configurations as follows:

1. Log in to the Data Warehouse service as an DWAdmin.
2. Go to Database Catalog ⋮ Edit CONFIGURATIONS Hue query processor and select hue-query-processor.json from the Configuration files drop-down menu.
3. Note the paths of the tez.history.logging.proto-base-dir and hive.hook.proto.base-directory properties.
4. After you upgrade the Database Catalog, go to Database Catalog ⋮ Edit CONFIGURATIONS Hue query processor and select hue-query-processor.json from the Configuration files drop-down menu.
5. Specify the desired path under the dasConf section as shown in the following example:

```
"dasConf": {
  "hive.hook.proto.base-directory": "/warehouse/tablespace/
external/hive/sys.db/query_data",
  "tez.history.logging.proto-base-dir": "/warehouse/tablespa
ce/external/hive/sys.db"
},
```

6. Click Apply Changes.

Known issues on OpenShift cluster environments

This topic describes the OpenShift cluster environment known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

Known issues identified before 1.4.1
DWX-4723: HTTP 504 Gateway timeout error returned when you delete an environment

When a CDW Private Cloud environment is deleted in the CDW Private Cloud UI, an HTTP 504 Gateway timeout error can be returned.

Using the OpenShift CLI, increase the route timeout to 60s on the OpenShift cluster:

```
oc annotate route [***ROUTE-NAME***] --overwrite haproxy.router.
openshift.io/timeout=60s
```

For more information about setting this timeout, see the OpenShift documentation.

ECS cluster environments

This topic describes the Embedded Container Service (ECS) cluster environment known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

BLESC-6074: Impala executor pods not running and queries from Hue fail

You may see the following error after submitting a query from Hue: Latest admission queue reason: Waiting for executors to start. Only DDL queries and queries scheduled only on the coordinator (either NUM_NODES set to 1 or when small query optimization is triggered) can currently run. You may also notice that the Impala executors are not running.

Make sure that the /etc/resolv.conf file on the ECS hosts contain a maximum of 2 search domains.

Known issues in Database Catalogs

This topic describes the Database Catalog known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

Known issues identified in 1.5.1
DWX-15302: Upgrade button stays visible even after the upgrade completes

After you upgrade the Database Catalog, the Upgrade button remains visible on the CDW web interface instead of disappearing or getting disabled.

Refresh the page on your browser.

Known issues in Hive Virtual Warehouses

This topic describes the Hive Virtual Warehouse known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.2
DWX-16989: Hive query running on Iceberg table fails randomly

Suppose you have disabled the auto-suspend option for a Hive Virtual Warehouse or if the Virtual Warehouse is under continuous load and hence it cannot be stopped by the auto-suspend option. In this situation, using the Iceberg table format may cause the following exceptions to appear in the query coordinator log along with the submitted queries that have failed:

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretMan
$InvalidToken): token (token for hive: HDFS_DELEGATION_TOKEN
 owner=hive/dwx-env-host-1.cdp.local@EXAMPLE.CLOUDERA.COM,
 renewer=hive, realUser=, issueDate=1709813340891,
 maxDate=1710418140891, sequenceNumber=19784486, masterKeyId=52)
 is expired, current time: 2024-03-08 04:09:32,835-0800 expected
 renewal time: 2024-03-08 04:09:00,891-0800
```

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretMan
$InvalidToken): token (token for hive: HDFS_DELEGATION_TOKEN
 owner=hive/dwx-env-host-1.cdp.local@EXAMPLE.CLOUDERA.COM,
 renewer=hive, realUser=, issueDate=1699855596578,
 maxDate=1700460396578, sequenceNumber=16863242, masterKeyId=39)
 can't be found in cache
```

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretMan
$InvalidToken): token (OzoneToken owner=hive/dwx-env-
ewxf6g-env.cdp.local@ROOT.EXAMPLE.SITE, renewer=hive,
 realUser=, issueDate=2024-03-19T21:49:31.033Z,
 maxDate=2024-03-19T21:50:31.033Z, sequenceNumber=72,
 masterKeyId=1, strToSign=null, signature=null,
```

```
awsAccessKeyId=null, omServiceId=ozone1710521984,
omCertSerialId=11) is expired, current time: 2024-03-19
21:51:34,293+0000 expected renewal time: 2024-03-19
21:51:31,033+0000
```

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretMan
$InvalidToken): token (OzoneToken owner=hive/dwx-env-
azt1gg-env.cdp.local@ROOT.EXAMPLE.SITE, renewer=hive,
 realUser=, issueDate=2024-04-09T16:04:12.889Z,
 maxDate=2024-04-09T17:04:12.889Z, sequenceNumber=29,
 masterKeyId=1, strToSign=null, signature=null,
 awsAccessKeyId=null, omServiceId=ozone1711550158,
 omCertSerialId=2597525731772327) can't be found in cache
```

This happens because the HDFS delegation tokens are not renewed when using the Iceberg table
format. After the existing HDFS delegation tokens expire, Hive query coordinator (TEZ App
Master) cannot access the tables on the file system during the query planning phase. The problem
is independent of the file system--Ozone FS or Hadoop FS. The error only occurs after the HDFS
delegation tokens have expired. By default, the delegation tokens expire in one day. However, you
can modify the expiration time on the CDP Base cluster.

The problem does not occur if the query coordinator pods in the Hive Virtual Warehouse are
stopped manually or by using the auto-suspend functionality within the token expiration period.

Apply this workaround only if you cannot suspend the Hive Virtual Warehouse.

1. Log in to the Data Warehouse service as DWAdmin.
2. Go to the Virtual Warehouses tab and click ⋮ Edit Configurations Query Coordinator .
3. Select env from the Configuration files drop-down menu.
4. Add the following value against the JVM_OPTS property:

```
-Diceberg.scan.plan-in-worker-pool=false
```

5. Click Apply Changes.

Known issues identified in 1.5.1

DWX-16891: Hive-Ranger integration issue after a refresh

You may see the following error after you refresh a Hive Virtual Warehouse: Error while compiling
statement: FAILED: HiveAccessControlException Permission denied. This happens because Hive
fails to evaluate Ranger group policies when the Virtual Warehouse is updated--either by upgrading
or refreshing it.

Rebuild the Hive Virtual Warehouse to fix the Ranger integration issues.

DWX-15480: Hive queries fail with FILE_NOT_FOUND error

ACID directory cache may become outdated in Tez AMs in case of ACID tables that change often,
possibly leading to different errors with the same root cause: "split generation works from cache
pointing to non-existing files". And you may see the following error in the diagnostic bundles and
query logs: FILE_NOT_FOUND: Unable to get file status.

Disable the cache by setting the value of the hive.txn.acid.dir.cache.duration property to -1 by going
to Virtual Warehouse ⋮ Edit CONFIGURATIONS Hue Configuration files hive-site from the
CDW web interface.

DWX-15287: Drop database query for Hive fails with Invalid ACL Exception

You may see the following error in a Hue or beeline session when running DROP DATABASE,
DROP TABLE, or ALTER TABLE DROP PARTITION operations on a Hive Virtual Warehouse

that is in Stopped state: "org.apache.zookeeper.KeeperException$InvalidACLException: KeeperErrorCode = InvalidACL for /llap-sasl/user-hive".

The exception is caused because the Hive Virtual Warehouse tries to evict the cache in the LLAP executors, but the compute pods of the stopped warehouse are no longer running.

> **Note:** The database or table is deleted despite the exception, only the LLAP executors do not flush their database or table related buffers, because these executors are not running.

Start the Virtual Warehouse before you run the DROP DATABASE, DROP TABLE, or ALTER TABLE DROP PARTITION operations.

Alternatively, you can add the hive.llap.io.proactive.eviction.enabled=false setting in the hive-sit e.xml file. This method may result in some performance degradation, because LLAP no longer discards the dropped database/table or temp table related buffers.

1. Log in to CDW as DWAdmin.
2. Click ⋮ Edit CONFIGURATIONS Hiveserver2 on the Virtual Warehouse tile and select hive-site from the Configuration files drop-down menu.
3. Click + and add the following line:

```
hive.llap.io.proactive.eviction.enabled=false
```

4. Click Apply Changes.

   Wait for the Virtual Warehouse to refresh and return to Running or Stopped state.

Known issues identified before 1.4.1

DWX-4842: Entities are not being created in Atlas

Base clusters that are using Java 11 might be using truststores in PKCS12 format. Currently, Hive Virtual Warehouses on CDW Private Cloud only supports truststores in JKS format. This prevents the entities from being created in Atlas.

Using the keytool, convert the PKCS12 truststore in the base cluster to a JKS truststore.

Known issues in Impala Virtual Warehouses

This topic describes the Impala Virtual Warehouse known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

Known issues identified in 1.5.1

DWX-14292: Impala executors and coordinator pods get stuck in pending state

If you have enabled low resource mode while activating an environment in CDW and also enabled the option to customize pod sizes for Impala from the **Advanced Settings** page, then you may notice that the executors and coordinator pods do not request the set amount of memory and CPU. This happens because the low resource mode overrides the default pod configuration.

Low resource mode and custom pod configurations for Impala cannot be used at the same time. To resolve this issue, do not select the Low resource mode option while activating an environment. Instead, create custom pod configurations with lower resources for Impala. You can continue to use lesser resources as defined in the custom pod configuration for Hive.

VIZ-2046: Trusted impersonation does not work in CDW

When you try to create a connection to the Impala Virtual Warehouse from Cloudera Data Visualization (CDV), CDV is unable to use Trusted Impersonation as a mode of authentication. Therefore, you need to enter both the username and password for authentication.

None.

Known issues in Hue

This topic describes the Hue known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.2

DWX-15100: Impala doesn't release the query profiles until the session is closed by the client

You may notice that the queries are stuck in the Waiting to be closed state and that the last query submitted from Hue is not displayed on the **Job Browse** page in Hue. Ideally, Impala must finish running the and the client should be able to fetch the results, and Hue should close the query, so that Impala can write the query profile to disk in thrift. However, this does not happen.

You can view the queries in one of the following ways:

- You can click on the query ID that is displayed under the query editor when a query starts to running and view the query profile.
- You can go to the **Job Browser** page's **Impala** tab to view the Impala query profile.
- You can close the Hue session to trigger query closure by clicking ⋮ Session Close .

Known issues identified in 1.5.1

CDPD-54376: Clicking the home button on the File Browser page redirects to HDFS user directory

When you are previewing a file on any supported filesystem, such as S3, ABFS, or Ozone and you click on the Home button, you are redirected to the HDFS user home directory instead of the user home directory on the said filesystem.

None.

DWX-15090: Intermittently see the CSRF error in the Hue Job Browser

You may intermittently see the "403 - CSRF" error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the page or start a new Hue session.

DWX-14911: Export operation failing in a partitioned Hive table

You may see the following error in the Hive server logs when running the EXPORT TABLE query in Hive: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.ExportTask. Distcp operation failed. This issue occurs when the number of files and the size of files exceeds the limits specified in the hive.exec.copyfile.maxsize and hive.exec.copyfile.maxnumfiles properties respectively, and the EXPORT TABLE query launches the distributed copy (`distcp`) job to copy files of HDFS.

Increase the limit of following properties in HiveServer2:

- hive.exec.copyfile.maxsize to maximum size of files in a directory in bytes
- hive.exec.copyfile.maxnumfiles to maximum number of files in a directory

OPSAPS-66903: Ozone HTTPFS address configuration contains placeholders instead of the real values

The Hue File Browser does not automatically support browsing the Ozone filesystem because the HTTPFS role configuration for the Ozone service contains placeholder text instead of real values. You can view this configuration by going to Cloudera Manager Ozone service Instances HTTPFS Gateway Processes and clicking on the ozone-conf/httpfs-site.xml configuration file.

You must manually configure Hue to access Ozone using the File Browser by following the instructions listed in Enabling browsing Ozone from Hue on CDW Private Cloud.

Known issues identified in 1.5.0

DWX-12616: Hue limitation in CDW Private Cloud

Following are the known limitations in Hue in CDW Private Cloud 1.5.0:

- Hue Importer does not support importing files more than 200 KB in size
- Hue File Browser does not work if the HDFS service on the base cluster is configured for high availability

None.

DWX-13865: Hue File Browser does not work with HDFS HA

Hue File Browser is not accessible or displays a 403 error when you click on File Browser from the left assist panel in CDW Private Cloud if HDFS is configured for High Availability on the base cluster. Currently, Hue in CDW cannot obtain the hostname and the port from the HttpFS service on the base cluster. This is a known limitation.

You must manually build and specify the WebHDFS URL for Hue in CDW to connect to the HttpFS service on the base cluster.

1. Log in to Cloudera Manager as an Administrator.
2. Go to Clusters Hive Configuration and note the value present in the Kerberos Principal field.

   This is the Hive service account name.
3. Go to Clusters HDFS HttpFS Advanced Configuration Snippet (Safety Valve) for httpfs-

   site.xml and click **+** to add the following lines:

   Name: httpfs.proxyuser.hive.hosts, Value: *

   Name: httpfs.proxyuser.hive.groups, Value: *

   Replace hive with the actual Hive service account name.
4. Click Save Changes and restart the HDFS service or the HttpFS role.
5. Go to the Instances tab and note the hostname of the HttpFS role.
6. Go to the Configuration tab and note the port for the hdfs.httpfs.http.port configuration from the RESTPort field.

   The default value of the hdfs.httpfs.http.port configuration is 14000.
7. Use the hostname and the port to construct the WebHDFS URL as follows:

   ```
   https://[***HOSTNAME***]:[***PORT***]/webhdfs/v1
   ```

8. Log in to the Data Warehouse service as a DWAdmin.
9. Go to the Virtual Warehouse from which you want to connect Hue to the base cluster's HDFS

   service and click **⋮** Edit .
10. Go to CONFIGURATIONS Hue , select hue-safety-valve from the Configuration files drop-down list and add the following lines:

    ```
    [hadoop]
    [[hdfs_clusters]]
    [[[default]]]
    webhdfs_url=https://[***HOSTNAME***]:[***PORT***]/webhdfs/v1
    ```

    Specify the WebHDFS URL that you constructed earlier.
11. Click Apply changes.

    Update the hue-safety-valve configuration for any Hive or Impala Virtual Warehouses from which you want to connect to the base cluster HDFS.

Known issues identified before 1.4.1

DWX-9373: Unable to log into Hue as a local administrator

> If you have logged into the CDP Management Console as a local administrator, then you may not be able to log into Hue and you may see the following error: "User is not authorized".
>
> To access Hue, you must add your username to your organization's LDAP tree or log into the CDP Management Console using your LDAP credentials and then access Hue.

Known issues in Unified Analytics

This topic describes the Unified Analytics known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.2

No new known issues identified in 1.5.2.

Iceberg-related known issues in CDW Private Cloud

This topic describes the Iceberg-related known issues in Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.2

DWX-16591: Concurrent merge and update Iceberg queries are failing

> You may see that the concurrent merge and update Iceberg queries fail with the following error in the Hive application logs: "Base metadata location hdfs://*<Location-A>* is not same as the current table metadata location '*<Location-B>*' for default.merge_insert_target_iceberg \rorg.apache.iceberg.exceptions.CommitFailedException". This happens because the corresponding Query A and Query B have overlapping updates. For example, if Query A commits the data and deletes files first, then Query B will fail with validation failure due to conflicting writes. In this case, Query B should invalidate the commit files that are already generated and re-execute the full query on the latest snapshot.
>
> None.

CDPD-59413: Unable to view Iceberg table metadata in Atlas

> You may see the following exception in the Atlas application logs when you create an Iceberg table from the CDW data service associated with a CDP Private Cloud Base 7.1.8 or 7.1.7 SP2 cluster: Type ENTITY with name iceberg_table does not exist. This happens because the Atlas server on CDP Private Cloud Base 7.1.8 and 7.1.7 SP2 does not contain the necessary, compatible functionality to support Iceberg tables. This neither affects creating, querying, or modifying of Iceberg tables using CDW nor does it affect creating of policies in Ranger.
>
> On CDP Private Cloud Base 7.1.9, Iceberg table entities are not created in Atlas. You can ignore the following error appearing in the Atlas application logs: ERROR - [NotificationHookConsumer thread-1:] ~ graph rollback due to exception (GraphTransactionInterceptor:200) org.apache.atlas.exception.AtlasBaseException: invalid relationshipDef: hive_table_storagedesc: end type 1: hive_storagedesc, end type 2: iceberg_table
>
> If you are on CDP Private Cloud Base 7.1.7 SP2 or 7.1.8, then you can manually upload the Iceberg model file z1130-iceberg_table_model.json in to the /opt/cloudera/parcels/CDH/lib/atlas/models/1000-Hadoop directory as follows:
>
> 1. SSH into the Atlas server host as an Administrator.
> 2. Change directory to the following:
>
> ```
> cd /opt/cloudera/parcels/CDH/lib/atlas/models/1000-Hadoop
> ```

3. Create a file called 1130-iceberg_table_model.json with the following content:

```json
{
  "enumDefs": [],
  "structDefs": [],
  "classificationDefs": [],
  "entityDefs": [
    {
      "name": "iceberg_table",
      "superTypes": [
        "hive_table"
      ],
      "serviceType": "hive",
      "typeVersion": "1.0",
      "attributeDefs": [
        {
          "name": "partitionSpec",
          "typeName": "array<string>",
          "cardinality": "SET",
          "isIndexable": false,
          "isOptional": true,
          "isUnique": false
        }
      ]
    },
    {
      "name": "iceberg_column",
      "superTypes": [
        "hive_column"
      ],
      "serviceType": "hive",
      "typeVersion": "1.0"
    }
  ],
  "relationshipDefs": [
    {
      "name": "iceberg_table_columns",
      "serviceType": "hive",
      "typeVersion": "1.0",
      "relationshipCategory": "COMPOSITION",
      "relationshipLabel": "__iceberg_table.columns",
      "endDef1": {
        "type": "iceberg_table",
        "name": "columns",
        "isContainer": true,
        "cardinality": "SET",
        "isLegacyAttribute": true
      },
      "endDef2": {
        "type": "iceberg_column",
        "name": "table",
        "isContainer": false,
        "cardinality": "SINGLE",
        "isLegacyAttribute": true
      },
      "propagateTags": "NONE"
    }
  ]
}
```

4. Save the file and exit.

5. Restart the Atlas service using Cloudera Manager.

Fixed issues in Cloudera Data Warehouse Private Cloud

This section lists issues fixed in this release of the Cloudera Data Warehouse (CDW) service.

DWX-14281: Configuration changes to Database Catalogs are not applied

Earlier, when you tried to change Database Catalog configurations from the CDW UI, the changes were not applied. This issue has been fixed.

DWX-15869: Database Catalog edit page does not load

Earlier, if the Trusted Kerberos Realms configuration on the base cluster's Core Settings service was non-empty, or if the Additional Rules to Map Kerberos Principals to Short Names field (auth_to_local_rule property) contained the following characters: \, ", \t, \n, \b, and \f, then the CDW Database Catalog and Virtual Warehouse edit page did not load. This issue has been fixed.

DWX-13362: Queries containing simple SELECT statements get cached

Earlier, a query such as SELECT * FROM planes LIMIT   2; used to get cached. This issue has been fixed.

DWX-15299: Atlas lineage does not work after upgrading from 1.5.0 to 1.5.1

This issue has been fixed.

DWX-15176: Hue frontend pods crashloop after upgrading from 1.5.0 to 1.5.1 on ECS

This issue has been fixed.

Version information for Cloudera Data Warehouse Private Cloud components

Cloudera Data Warehouse (CDW) uses Hive, Impala, and Hue as its Runtime components and also provides integration with Cloudera Data Visualization. Review the version information of CDW Private Cloud 1.5.2 components.

Table 1: CDW Private Cloud version information

CDW component	Version
Hive	2023.0.15.2
Impala	2023.0.15.2
Hue	2023.0.15.2
Cloudera Data Visualization	7.1.6
CDW server	1.8.0
CDP CLI	0.9.101 to 0.9.107