Cloudera Data Warehouse Private Cloud 1.5.1

Data Warehouse Release Notes

Date published: 2020-08-17
Date modified: 2023-06-13

CLOUD≡RA

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 1.5.1.

Ability to upgrade Database Catalogs and Virtual Warehouses

After upgrading the CDP platform to the latest version, you can upgrade the Database Catalogs and Virtual Warehouses to the latest available versions from the CDW web interface or from the CLI. The option to upgrade is displayed only if you are upgrading CDP Private Cloud platform from 1.5.0 to a newer version (1.5.1). For more information, see Upgrading Database Catalogs and Virtual Warehouses in CDW Private Cloud.

CDW supports Kerberos authentication for Beeline, Impala-shell, Impyla, and other JDBC clients

In addition to LDAP, you can now use Kerberos to authenticate requests from CLI-based Business Intelligence clients such as Beeline, Impala-shell, Impyla, and other JDBC clients. You can store user credentials in Kerberos Keytab files, so that you can run commands without having to specify the username and password as command-line parameter or enter them after running a command. See Authenticating users in CDW Private Cloud.

Group-level access control for Impala Virtual Warehouse

CDW enables you to allow one or more user groups to access a particular Impala Virtual Warehouse. As a result, only specific users can connect to a Virtual Warehouse, from all supported channels (Hue, JDBC, or other Business Intelligence tools). The option to specify user groups when creating a new Virtual Warehouse is enabled by default. You can disable this option from the **Advanced Settings** page in the Data Warehouse UI. See Configuring access control for Impala Warehouses.

Enhancements to the Impala custom pod configuration UI

The user experience to create and customize the custom pod configurations for Impala has been enhanced. See the revised how-to topic Creating custom pod configurations for Impala Virtual Warehouses.

Workload-aware autoscaling for Impala (Preview)

Using workload-aware autoccaling, you can configure multiple executor groups within a single Virtual Warehouse that can independently autoscale to allow handling of different workloads in the same Virtual Warehouse. According to each query's resource requirement, the query is scheduled on an executor group size that is appropriate for that query. For more information, see Workload aware autoscaling in Impala.

> **Note:** This feature is in technical preview and not recommended for production deployments. Cloudera recommends that you try this feature in development or test environments.

You must select the Enable workload-aware autoscaling for Impala option from the Advanced Configurations to use workload-aware autoscaling. See Enabling workload-aware autoscaling for Impala.

Configurations are copied from base to CDW for seamless workload migration

Configurations such as default table types, compression format, file format, parquet resolution, timezones, and so on are copied from CDP Private Cloud Base to CDW Data Service when you activate an Environment in CDW. This enables seamless migration of Impala and Hive LLAP workloads to CDW. See List of configurations copied from the Base cluster to CDW Private Cloud.

Ability to browse the Ozone filesystem using the Hue File Browser

You can browse files and directories on the Ozone filesystem on the base cluster just like you can on S3 or ADLS Gen2. To enable browsing files on Ozone, you must first enable the Ozone File Browser. See Enabling browsing Ozone from Hue.

Support for Ozone erasure-coded data in Impala

Impala now supports reading from Ozone data stored with Erasure Coding (EC). The Ozone EC feature provides data durability and fault tolerance along with reduced storage space and ensures data durability similar to the Ratis THREE replication approach. EC can be considered as an alternative to replication. For more information, see Erasure Coding Overview.

Support for Ozone erasure-coded data in Hive

Hive now supports reading data stored in Ozone that is enabled with Erasure Coding (EC). The Ozone EC feature provides data durability and fault-tolerance along with reduced storage space and ensures data durability similar to Ratis/THREE replication approach. For more information, see Using Ozone storage with Cloudera Data Warehouse.

The "enable_queries_list" configuration has been removed from Hue jobbrowser safety valve section

The enable_queries_list configuration in the CDW Hue safety valve was used to display or hide the **Queries** tab on the **Job Browser** page. This configuration has been removed. The **Queries** tab is displayed by default based on the type of Virtual Warehouse, whether it is Hive or Impala. You can override the query_store   configuration and hide the **Queries** tab. For more information, see Hue configurations in Cloudera Data Warehouse.

Hive ACID compaction observability

Compaction observability is a notification and information system based on metrics about the health of the compaction process. You can access Grafana dashboards from CDW web interface to view alerts about compaction status, the issue, and recommended actions.

The following list describes a few of the 25 notifications:

- Oldest initiated compaction passed threshold
- Large number of compaction failures
- More than one host is initiating compaction

Create an environment in this release to use this feature. For more information, see Compaction observability.

CDW in Private Cloud supports Apache Iceberg V2 on HDFS and Ozone (Preview)

Apache Iceberg integration with Cloudera Data Platform (CDP) enhances the Lakehouse architecture by extending multifunction analytics to a petabyte scale for multi-cloud and hybrid use cases. From Hive or Impala, you use Apache Iceberg features in CDW on HDFS and Ozone, which include time travel, create table as select (CTAS), and schema and partition evolution. To use Iceberg with CDW, you must upgrade to CDP Private Cloud Data Services 1.5.1.

Apache Iceberg V1 and V2 is in technical preview in the Private Cloud 1.5.1 release and is not recommended for production deployments. Cloudera recommends that you try this feature in test or development environments.

Note: Iceberg tables cannot be accessed from Hive, Impala, Spark, and Flink on the base cluster. You need to be on CDW. For the list of supported features and more information, see Iceberg overview.

Deprecations

Data Analytics Studio (DAS) has been deprecated and is no longer available with CDW Private Cloud starting with 1.5.1. DAS features have been migrated to the Hue Query Processor, along with the ability to query data in Hive, Impala, and Unified Analytics mode. After you upgrade to this release, any DAS instances that are running will be removed from the cluster.

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 1.5.1

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 in 1.5.0

DWX-18903: Service "postgres-service-default-warehouse" is invalid: spec.externalName error

> You see the following error during the Database Catalog creation stage after activating the environment in CDW:

```
Service "postgres-service-default-warehouse" is invalid:
 spec.externalName
a lowercase RFC 1123 subdomain must consist of lower case
 alphanumeric characters, '-' or '.', and must start and end with
 an alphanumeric character (e.g. 'example.com', regex used for
 validation is '[a-z0-9]([-a-z0-9]*[a-z0-9])?(\.[a-z0-9]([-a-
z0-9]*[a-z0-9])?)*')
```

> . This could happen because if the value of the Hive Metastore Database Host (hive_metastore_database_host) property on the base cluster is not specified in lowercase.
>
> Go to Cloudera Manager Clusters Hive service Configuration and change the value specified in the Hive Metastore Database Host field to be in lowercase.

Known Issues identified before 1.4.1

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

> If you upgrade the Database Catalogs and the Virtual Warehouses after upgrading the platform from 1.5.0 to 1.5.1, then the Kafka producers have idempotence enabled by default. Idempotent producers

must have idempotent write permission set on the cluster resource in Ranger. If idempotence is not disabled or if the write permission is not set correctly, then Atlas integration fails with the following error Entity is not found in atlas, please check metastore pod/atlas application logs'.

To resolve this issue, you can either:

- Explicitly disable idempotence by setting the value of the atlas.kafka.enable.idempotence property to false in hive-atlas-config for Metastore, Hiveserver2, Hive query-executor and Impala coordinator.
- Update your policies in Ranger and ensure that Hive and Impala users have Idempotent Write permission on the cluster resource.

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

After you upgrade from 1.5.0 to 1.5.1 on Embedded Container Service (ECS), Virtual Warehouses go into an error state as Hue frontend pods crash. This is due to a breaking change in the underlying Kubernetes version.

Upgrade the Virtual Warehouses to upgrade their runtime software by clicking ⋮ Upgrade .

DWX-8525 and DWX-8526: Known limitations for upgrading from 1.4.1. to 1.5.1

Hue does not work after upgrade due to DWX-15176 (Hue frontend pods crashloop) after upgrading from 1.4.1 to 1.5.1.

You must reactivate the CDW environment after upgrading the Private Cloud data services from 1.4.1 to 1.5.1.

> ⚠ **Important:** Review the following caveats before upgrading from 1.4.1 to 1.5.1:
>
> - Reactivating the environment deletes the existing Database Catalogs and Virtual Warehouses. However, the metadata stored in the Hive Metastore (HMS) and data that is stored on HDFS or Ozone from the previous version is not deleted from the base cluster. Cloudera recommends that you manually take a backup of the Virtual Warehouse configurations, which you can use to create new Virtual Warehouses after the upgrade.
> - Apply workaround for the AD limitation (TSB-639) on the base cluster. Cloudera recommends that you do not use deprecated encryption types in production environments. Use stronger encryption types such as AES 256 and AES 128. For more information, see Resolving Kerberos authentication failure.

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 1.5.1

No new known issues in this area.

ECS cluster environments

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

Known Issues identified before 1.4.1
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 1.5.1
DWX-15869: Database Catalog edit page does not load

If the "Trusted Kerberos Realms" configuration on the base cluster's Core Settings service is not empty or if the Additional Rules to Map Kerberos Principals to Short Names field (auth_to_local_ru le property) contains the following characters: \, ", \t, \n, \b, and \f, then the CDW Database Catalog and Virtual Warehouse edit page does not load. The logs of the dwx-server pod in the control plane namespace contain the following error message: Not able to render config, err:invalid character 'Q' in string escape code.

If the Trusted Kerberos Realms property is configured, then perform the folowing:

1. Log in to Cloudera Manager as an Administrator.

2. Go to Clusters Core Settings service Actions , click Download Client Configuration. Save the file on your computer and decompress it.

3. Open the core-site.xml file and copy the value of the hadoop.security.auth_to_local configurat ion property.

4. Paste the value of the hadoop.security.auth_to_local configuration property in a text editor and remove all \Q and \E regular expression constructs. Copy the modified text.

5. Go to Cloudera Manager Clusters Core Settings service Configuration and paste the modified text in the Additional Rules to Map Kerberos Principals to Short Names field.

6. Remove the Trusted Kerberos Realms configuration from Core Settings service Configuration tab.

7. Click Save Changes.

8. Restart the affected services in Cloudera Manager to apply stale configurations.

9. Go to the Data Warehouse service and refresh your Database Catalogs and Virtual Warehouses. In case this does not work, deactivate and reactivate your environment in CDW.

If the Additional Rules to Map Kerberos Principals to Short Names field (extra_auth_to_local_ rules property) in Core Settings service contains the following characters: \, ", \t, \n, \b, and \f, then remove these characters regular expression. Save changes and restart the affected services in Cloudera Manager, and then refresh the Database Catalog and Virtual Warehouses.

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 identified in 1.5.0
DWX-14022: The metastore-sys-init-job job fails with "Table not found 'TXNS'" error

You see that the default Database Catalog starts successfully and the UI indicates it as running, however, the metastore-sys-init pods are in an error state. You also see that the Hive MetaStore (HMS) pods fail with the following error: "[HiveServer2-Background-Pool: Thread-125] ERROR org.apache.hadoop.hive.ql.parse.CalcitePlanner - org.apache.hadoop.hive.ql.parse.SemanticException: Line 1:1117 Table not found 'TXNS'". This happens because the TXNS table is not present on CDP Private Cloud Base version 7.1.7 SP2.

You must upgrade the base cluster to CDP Private Cloud Base version 7.1.8 and then rebuild the affected Database Catalog.

Known Issues identified in 1.4.1
DWX-13758: Updating the pre-upgrade Database Catalog and Virtual Warehouse fails silently

If you update or refresh the Database Catalogs and Virtual Warehouses that were created before you upgraded to 1.4.1 by changing their configurations, then they may go into an erroneous state.

Note: This issue is true if you are upgrading from 1.4.1 to 1.5.1. If you are upgrading from 1.5.0, you would not face this issue.

Do not update or refresh the Database Catalogs and Virtual Warehouses that were created before you upgraded to 1.4.1.

Create new Database Catalogs and Virtual Warehouses.

Known Issues identified before 1.4.1
DWX-8979: Parquet demo data set only loads partially on OpenShift clusters

On OpenShift clusters, the Parquet demo data set loads partially on a non-default Database Catalog, intermittently. As a result, you may intermittently and partially see the demo data set in Hue.

None.

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 1.5.1

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-15869: Virtual Warehouse edit page does not load

If the "Trusted Kerberos Realms" configuration on the base cluster's Core Settings service is not empty or if the Additional Rules to Map Kerberos Principals to Short Names field (auth_to_local_ru le property) contains the following characters: \, ", \t, \n, \b, and \f, then the CDW Database Catalog and Virtual Warehouse edit page does not load. The logs of the dwx-server pod in the control plane namespace contain the following error message: Not able to render config, err:invalid character 'Q' in string escape code.

If the Trusted Kerberos Realms property is configured, then perform the folowing:

1. Log in to Cloudera Manager as an Administrator.
2. Go to Clusters Core Settings service Actions , click Download Client Configuration. Save the file on your computer and decompress it.
3. Open the core-site.xml file and copy the value of the hadoop.security.auth_to_local configurat ion property.
4. Paste the value of the hadoop.security.auth_to_local configuration property in a text editor and remove all \Q and \E regular expression constructs. Copy the modified text.
5. Go to Cloudera Manager Clusters Core Settings service Configuration and paste the modified text in the Additional Rules to Map Kerberos Principals to Short Names field.
6. Remove the Trusted Kerberos Realms configuration from Core Settings service Configuration tab.
7. Click Save Changes.
8. Restart the affected services in Cloudera Manager to apply stale configurations.
9. Go to the Data Warehouse service and refresh your Database Catalogs and Virtual Warehouses. In case this does not work, deactivate and reactivate your environment in CDW.

If the Additional Rules to Map Kerberos Principals to Short Names field (extra_auth_to_local_ rules property) in Core Settings service contains the following characters: \, ", \t, \n, \b, and \f, then remove these characters regular expression. Save changes and restart the affected services in Cloudera Manager, and then refresh the Database Catalog and Virtual Warehouses.

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 1.5.1

DWX-15182: Performance degradation due to a configurations copied from the base cluster

If you have enabled the Copy configurations from base cluster to CDW option from Advanced Configurations, then all base cluster configurations, including "fe_service_threads" are copied to CDW. The fe_service_threads configuration impacts client concurrent connections. You may see degraded performance if the value of this configuration is set to anything less than 96 on the base cluster.

Ensure that the value of this property is at least 96.

DWX-15869: Virtual Warehouse edit page does not load

If the "Trusted Kerberos Realms" configuration on the base cluster's Core Settings service is not empty or if the Additional Rules to Map Kerberos Principals to Short Names field (auth_to_local_ru le property) contains the following characters: \, ", \t, \n, \b, and \f, then the CDW Database Catalog and Virtual Warehouse edit page does not load. The logs of the dwx-server pod in the control plane namespace contain the following error message: Not able to render config, err:invalid character 'Q' in string escape code.

If the Trusted Kerberos Realms property is configured, then perform the folowing:

1. Log in to Cloudera Manager as an Administrator.
2. Go to Clusters Core Settings service Actions , click Download Client Configuration. Save the file on your computer and decompress it.
3. Open the core-site.xml file and copy the value of the hadoop.security.auth_to_local configurat ion property.
4. Paste the value of the hadoop.security.auth_to_local   configuration property in a text editor and remove all \Q and \E regular expression constructs. Copy the modified text.
5. Go to Cloudera Manager Clusters Core Settings service Configuration and paste the modified text in the Additional Rules to Map Kerberos Principals to Short Names field.
6. Remove the Trusted Kerberos Realms configuration from Core Settings service Configuration tab.
7. Click Save Changes.
8. Restart the affected services in Cloudera Manager to apply stale configurations.
9. Go to the Data Warehouse service and refresh your Database Catalogs and Virtual Warehouses. In case this does not work, deactivate and reactivate your environment in CDW.

If the Additional Rules to Map Kerberos Principals to Short Names field (extra_auth_to_local_ rules property) in Core Settings service contains the following characters: \, ", \t, \n, \b, and \f, then remove these characters regular expression. Save changes and restart the affected services in Cloudera Manager, and then refresh the Database Catalog and Virtual Warehouses.

DWX-15537: Kerberos authentication throwing "user not authorized error" after updating a usergroup in an Impala Virtual Warehouse

You may see an "User is not authorized" error when opening a session with Kerberos authentication to an Impala Virtual Warehouse where the access is restricted to user groups listed in the Virtual Warehouse settings. This problem may occur only when trying to connect with the same user as the delegation user specified during CDP Private Cloud Data Services environment activation.

To successfully connect to an Impala Virtual Warehouse from an impala-shell client using Kerberos, define the Impala delegation user (impala.doas.user) in the JDBC URL or in the impala-shell command. Ensure that the username is exactly the same as the one that you used to activate the Private Cloud Data Services environment and for which a Kerberos ticket was obtained at the time of the error.

DWX-15032: Impala Virtual Warehouse transitions to a running state way before the update finishes

When an Impala Virtual Warehouse is in an Updating or Upgrading state after an edit or upgrade operation the Virtual Warehouse transitions to a Running or Stopped state before the update process finishes. When the warehouse is in a Running or Stopped state after updating it, the changes made to the Virtual Warehouse, might not take effect until later.

To confirm whether the update has completed, check if the following log message is displayed in the dwx-server logs: "Virtual Warehouse update completed".

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 1.5.1

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 1.5.0

DWX-13362: Queries containing simple SELECT statements get cached

> A query such as SELECT * FROM planes LIMIT 2; gets cached. This is an undesirable result and will be fixed in a future release.

> None.

Iceberg known issues

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

Technical Service Bulletins

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

> Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

> - DELETE statements
>   - Deleting records pointing to old files have no effect
> - UPDATE / MERGE statements
>   - Deleting records pointing to old files have no effect
>   - The table will also have the newly added data records
>   - Rewritten records will still be active

> This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

> For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

> The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old "dangling" position delete files.

> Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

> The following conditions must be met for this issue to occur:

> - All delete files in the Iceberg table are "dangling"
>   - This would occur immediately after running Spark rewrite_data_files AND
>   - Before any further delete operations are performed on the table OR
>   - Before Spark rewrite_position_delete_files is run on the table

- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:

  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

  Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

Fixed issues in Cloudera Data Warehouse Private Cloud

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

OPSAPS-64734: Unable to add Impala as a new service if the default Hive warehouse directory is on Ozone

Earlier, when you tried to add the Impala service when the default Hive warehouse directory was configured to be on Ozone, the operation failed because o3fs and ofs filesystem interfaces do not support modifying the ACLs. This issue has been fixed.

OPSAPS-58019: "SERVICE_PRINCIPAL is required for kinit" error while activating a new environment

Earlier, if the /etc/krb5.conf file on the Cloudera Manager host contained "include" or "includedir" directives, then you encountered Kerberos-related failures on both Embedded Container Service and RedHat OpenShift platforms, and you had to comment the lines containing "include" and "includedir" directives in the /etc/krb5.conf file. This issue has been fixed, and you no longer need to comment these directives.

DWX-13753: Multiple CDW PVC Environments in same AD Domain (causes errors like Invalid Kerberos principals)

Earlier, when you created a Private Cloud cluster, the Control Plane created a default ECS environment with the name "default". Because CDW uses deterministic namespace names by default, the Kerberos keytabs on the existing ECS cluster were invalidated by values from the new ECS cluster's Cloudera Manager when you had multiple ECS clusters managed using different instances of Cloudera Manager but using the same Active Directory (AD) server. This caused the existing ECS environment to become unusable.

As a workaround, you had to disable the Use deterministic namespaces names option in the CDW **Advanced Settings** page before activating an environment in CDW.

This issue has been fixed. The default ECS environment is now created with a unique name by the Control Plane whenever you create a Private Cloud cluster. The environment name includes a portion of the ECS cluster's app domain. You no longer have to disable the Use deterministic namespaces names option in the CDW **Advanced Settings** page before activating an environment in CDW.

If you want to create ECS or OCP environments with the same name from the respective Cloudera Manager instances, then you must deselect the Use deterministic namespaces names option in the CDW **Advanced Settings** page before activating an environment in CDW. Otherwise, the older environments become unusable.

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

This issue has been fixed

DWX-14302: Hive query iolation executor pod crash loops with "Failed to start LLAP Daemon" exception

> Earlier, if you enabled the Query Isolation option on a Hive Virtual Warehouse, the queries that should run in isolation failed to start the executor, and eventually timed out with an exception. This issue has been fixed.

DWX-6234: WAIT TIME seconds cannot be set to "0" in the Hive Virtual Warehouse UI

> Earlier, when you set the WAIT TIME in the Hive Virtual Warehouse to 0, the UI automatically used to change it to 60. This issue has been fixed.

DWX-13934: Impala returns 401 Unauthorized error when connecting to an Impala Virtual Warehouse

> Earlier, Hue and Impala shell displayed a "401 Unauthorized error" when you submitted a query to an Impala Virtual Warehouse. This issue has been fixed.

DWX-8348: Service Discovery does not work when Impala is enabled with Unified Analytics

> This issue has been fixed.

DWX-13891: Hive queries intermittently fail with "InvalidACLException" error

> Earlier, you may have encountered the following error in Hue or Beeline when you run the DROP DATABASE, DROP TABLE, or ALTER TABLE DROP PARTITION operations on an Impala Virtual Warehouse running in the Unified Analytics mode: org.apache.zookeeper.KeeperException $InvalidACLException: KeeperErrorCode = InvalidACL for /llap-sasl/user-hive. This issue has been fixed.

DWX-14536: HiveServer2 in Unified Analytics Does Not Apply Ranger Security Policies

> In Cloudera Data Warehouse (CDW), HiveServer2 (HS2) in Unified Analytics is not using Rang erHiveAuthorizerFactory, causing it to ignore Ranger security policies due to incorrect code paths.

> This issue has been fixed.

Fixed issues in Cloudera Data Warehouse Private Cloud (1.5.1-CHF1)

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

DWX-15933: Make Atlas service as an optional dependency in CDW

> Apache Atlas is no longer a required dependency for deploying and using CDW. You can now deploy CDW without needing to install Atlas on the base cluster.

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.

Runtime component versions for Cloudera Data Warehouse Private Cloud

Cloudera Data Warehouse (CDW) uses Hive, Impala, and Hue as its Runtime components. The following table lists the versions of Hive, Impala, and Hue used in the CDW Private Cloud 1.5.1 release:

Table 1: CDW Runtime version information

Runtime component	Version
Hive	2023.0.14.1
Impala	2023.0.14.1
Hue	2023.0.14.1

Table 2: CDP CLI version information

CDW server version	CDP CLI version
1.7.0	Upto and including 0.9.100