Cloudera Data Warehouse on premises 1.5.5

Cloudera Data Warehouse Release Notes

Date published: 2020-08-17
Date modified: 2025-11-08

https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

What's new in Cloudera Data Warehouse on premises

Review the new features in this release of Cloudera Data Warehouse on premises service.

- Cloudera Data Warehouse
- Hive
- Impala
- Iceberg
- Hue

Cloudera Data Warehouse on premises

Integrating third-party Certification Manager

Cert-manager is an open-source tool for Kubernetes that automates the provisioning, management, and renewal of TLS certificates. Its documentation at https://cert-manager.io/docs/ provides comprehensive guidance on installing, configuring, and using cert-manager to secure workloads with trusted X.509 certificates. Cloudera provides out-of-the-box support for Venafi TPP as part of the Cloudera Embedded Container Service installation. By integrating cert-manager, the Cloudera Data Services on premises achieve secure communication, reduced manual overhead, and compliance with security standards, leveraging its robust automation and flexibility. For more information on integrating Cert-manager using Venafi TPP in Cloudera Data Warehouse, see Configuring cluster issuer for Certificate Manager.

Quota management improvements to support multiple environments

As part of this release, Quota Management capabilities have been enhanced to support multiple environments. Previously, root served as the top-level resource for the cluster. With the new changes, each environment now has its own resource pool for the respective data service.

When an environment is activated in Cloudera Data Warehouse, a root.<environment-name>.cdw resource pool is automatically created. This newly created resource pool can be selected as the top-level resource pool. For more details, refer to Quota management in Cloudera Data Warehouse on premises.

Improvements to Impala Autoscaler Dashboard

The following inprovements were introduced for the Impala Autoscaler Dashboard:

- Ability to select the log-level configuration for the autoscaler and autoscaler metrics containers.
- A new “Understanding The Dashboard” page has been added which explains the metrics displayed on the UI and how they are calculated.
- Empty data points that manifest as gaps in the graphs are skipped. Zero values are accurately displayed.

For more information, see About Impala Autoscaling dashboard.

Ability to view end-of-support information through UI and CDP CLI

Cloudera Data Warehouse releases reach the end of support every six months. The Cloudera Data Warehouse UI displays whether your deployment is nearing its end of support time or is unsupported, enabling you to plan an upgrade. You can also view the upgrade instructions on the UI. The end of support information is also displayed when you run the list-clusters and describe-clu sters CDP CLI commands.

Streamlined option for downloading Cloudera Data Warehouse diagnostic bundles

Cloudera Data Warehouse users can now easily download diagnostic bundles with a direct Collect option that reduces the need for prior time interval and log selection adjustments. This update enables faster, more efficient access to relevant diagnostic data. See, Downloading diagnostic bundles and Accessing and generating diagnostic bundles

Security improvement: use of Chainguard images

To enhance security, Cloudera Data Warehouse now uses Chainguard hardened images for its base images, Impala, Hue, and third-party images. The Kubernetes Dashboard is excluded from this change.

These changes help us address CVEs and offer improved security and stability. For more information, see Chainguard container images.

What's new in Hive on Cloudera Data Warehouse on premises

Hive Query History Service

The Hive query history service provides a scalable solution for storing and analyzing historical Hive query data. It captures detailed information about completed queries, such as runtime, accessed tables, errors, and metadata, and stores it in an efficient Iceberg table format. For more information see, Hive query history service

OpenTelemetry integration for Hive

Hive now integrates with OpenTelemetry (OTel) to enhance query by collecting and exporting telemetry data, including infrastructure and workload metrics. An OTel agent in Cloudera Data Warehouse helps monitor query performance and troubleshoot failures. For more information, see OpenTelemetry support for Hive

Apache Jira: HIVE-28504

What's new in Impala on Cloudera Data Warehouse on premises

Improved Cardinality Estimation for Aggregation Queries

Impala now provides more accurate cardinality estimates for aggregation queries by considering data distribution, predicates, and tuple tracing. Enhancements include:

- Pre-aggregation Cardinality Adjustments: A new estimation model accounts for duplicate keys across nodes, reducing underestimation errors.
- Predicate-Aware Cardinality Calculation: The planner now considers filtering conditions on group-by columns to refine cardinality estimates.
- Tuple Tracing for Better Accuracy: Improved tuple analysis allows deeper tracking across views and intermediate aggregation nodes.
- Consistent Aggregation Node Stats Computation: The planning process now ensures consistent and efficient recomputation of aggregation node statistics. These improvements lead to better memory estimates, optimized query execution, and more efficient resource utilization.
- Tuple-Based Cardinality Analysis: Analyzing grouping expressions from the same tuple to ensure their combined number of distinct values does not exceed the output cardinality of the source PlanNode, reducing overestimation.
- Refined number of distinct values Calculation for CPU Costing: The new approach applies a probabilistic formula to a single global NDV estimate, improving accuracy and reducing overestimation in processing cost calculations.

Apache Jira: IMPALA-2945, IMPALA-13086, IMPALA-13465 , IMPALA-13526, IMPALA-13405 IMPALA-13644

Cleanup of host-level remote scratch dir on startup and exit

Impala now removes leftover scratch files from remote storage during startup and shutdown, ensuring efficient storage management. The cleanup targets files in the host-specific directory within the configured remote scratch location.

A new flag, remote_scratch_cleanup_on_start_stop, controls this behavior. By default, cleanup is enabled, but you can disable it if multiple Impala daemons on a host or multiple clusters share the same remote scratch directory to prevent unintended deletions.

Apache Jira: IMPALA-13677, IMPALA-13798

Graceful shutdown with query cancellation

Impala now attempts to cancel running queries before reaching the graceful shutdown deadline, ensuring resources are released properly. The new shutdown_query_cancel_period_s flag controls this behavior. The default value is 60 seconds. If set to a value greater than 0, Impala will try to cancel running queries within this period before forcing shutdown. If the value exceeds 20% of the total shutdown deadline, it is automatically capped to prevent excessive delays. This approach helps prevent unfinished queries and unreleased resources during shutdown. For more information, see Setting Impala Query Cancellation on Shut down

Programmatic query termination

Impala now supports the KILL QUERY statement, enabling you to forcibly terminate queries for better workload management. The KILL QUERY statement cancels and unregisters queries on any coordinator. For more information, see KILL QUERY statement

Ability to log and manage Impala workloads

Cloudera Data Warehouse provides you the option to enable logging Impala queries on an existing Virtual Warehouse or while creating a new Impala Virtual Warehouse. The information for all completed Impala queries is stored in the sys.impala_query_log system table. Information about all actively running and recently completed Impala queries is stored in the sys.impala_query_live system table. Users with appropriate permissions can query this table using SQL to monitor and optimize the Impala engine. For more information, see Impala workload management

What's new in Iceberg on Cloudera Data Warehouse on premises

Cloudera support for Apache Iceberg version 1.5.2

The Apache Iceberg component has been upgraded from 1.4.3 to 1.5.2.

Reading Iceberg Puffin statistics

Impala supports reading Puffin statistics from current and older snapshots. When there are Puffin statistics for multiple snapshots, Impala chooses the most recent statistics for each column. This indicates that statistics for different columns may come from different snapshots. If there are Hive Metastore (HMS) and Puffin statistics for a column, the most recent statistics are considered. For HMS statistics, the impala.lastComputeStatsTime property is used and for Puffin statistics, the snapshot timestamp is used to determine which among the two is the most recent. For more information, see Iceberg Puffin statistics.

Note: Reading Puffin statistics is disabled by default. Set the --enable_reading_puffin_stats startup flag to "true" to enable it.

Enhancements to Iceberg data compaction

The OPTIMIZE TABLE statement is enhanced with the following improvements:

- **Supports partition evolution**

  The Hive and Impala OPTIMIZE TABLE statement that is used to compact Iceberg tables and optimize them for read operations, is enhanced to support compaction of Iceberg tables with partition evolution.
- **Supports data compaction based on file size threshold**

  The Impala OPTIMIZE TABLE statement has been enhanced to include a FILE_SIZE_THRESHOLD_MB option that enables you to specify the maximum size of files (in MB) that should be considered for compaction.

For more information, see Iceberg data compaction.

Impala supports the MERGE INTO statement for Iceberg tables

You can use Impala to run a MERGE INTO statement on an Iceberg table based on the results of a join between a target and source Iceberg table. For more information, see the Iceberg Merge feature.

What's new in Hue on Cloudera Data Warehouse on premises

Enhanced AI Integration in Hue SQL AI Assistant

The Hue SQL AI Assistant now supports Cloudera AI Workbench and Cloudera AI Inference service. These integrations enhance the Hue SQL AI Assistant by enabling the use of private models hosted within Cloudera-managed infrastructure. This ensures enhanced security and privacy while leveraging GenAI for the Hue SQL-related tasks.

- Cloudera AI Workbench: This enables you to securely deploy and run your own models within a virtual private cloud. This configuration enhances control and privacy within your environment. For more information, see Configure SQL AI Assistant using Cloudera AI Workbench.
- Cloudera AI Inference service: Helps in a production-grade serving environment for hosting predictive and generative AI models. This service simplifies model deployment and maintenance. For more information, see Configure SQL AI Assistant using Cloudera AI Inference service.

Hue SQL AI: Multi database querying now supported

The Hue SQL AI Assistant now supports multi-database querying, allowing you to retrieve data from multiple databases simultaneously. This enhancement simplifies managing large datasets across different systems and enables seamless cross-database queries.

- Support for cross-database queries.
- Ability to retrieve and combine data from multiple sources in a single query.

For more information, see Multi database support for SQL query.

User Input Validation for Hue SQL AI

Hue SQL AI now supports secure and optimized integration with large language models (LLMs). You can now configure user input validation, such as prompt length limits, regex restrictions, and HTML tag handling, and more to enhance both security and system performance.

For more information, see User Input Validation for Hue SQL AI.

Known issues and limitations in Cloudera Data Warehouse on premises

Review the known issues and limitations that you might run into while using the Cloudera Data Warehouse service in Cloudera Private Cloud Data Services.

General known issues

This topic describes the general service-wide known issues for Cloudera Data Warehouse on premises.

Known issues identified in 1.5.5

DWX-21003: Impala Virtual Warehouses with Quota Management enabled remain in the 'Pending' state

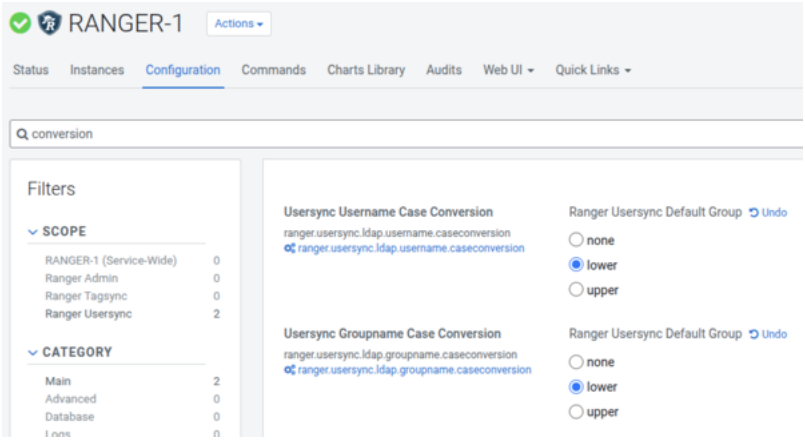
After an upgrade, Impala Virtual Warehouses with Quota Management enabled may fail to start and remain in the 'Pending' state due to insufficient quota resources.

If you are upgrading from Cloudera Data Warehouse on premises version 1.5.4 SP2 to 1.5.5, you can refresh, update, or upgrade the Impala Virtual Warehouse. From earlier versions, you can upgrade the Impala Virtual Warehouse.

Case sensitivity issue for Ranger authentication

In Active Directory environments, user and group names are often written in mixed case (for example, 'JohnDoe' or 'AdminGroup') and handled in a case-insensitive manner by Windows. However, Cloudera Base on premises operates in a Linux environment, where names are case-

sensitive. To address this, some customers configure Cloudera Base on premises to disable case sensitivity in System Security Service Daemon (sssd) and modify Ranger Usersync settings to convert user and group names to lowercase, ensuring compatibility with Ranger policies.



While this configuration works correctly in Cloudera Base on premises, authorization issues may arise inCloudera Data Warehouse components like Hive and Impala. Cloudera Data Warehouse does not automatically convert group names to lowercase, causing mismatches with Ranger policies that define group names in lowercase. This can result in authorization problems, such as users being unable to access databases, tables, or columns in Hue or remote client shells (impala-shell or jdbc), even though access works correctly in Cloudera Base on premises Hue or remote client shells.

To resolve this issue, enable group name conversion to lowercase in Cloudera Data Warehouse by adding the following Hadoop core-site configuration entries to the hadoop-core-site-default-warehouse configuration file. For Hive Virtual Warehouse, apply the changes to HiveServer2. For Impala Virtual Warehouse, apply the changes to Impala Catalogd, Impala Coordinator, Impala Executor, and Impala StateStored.

Property Name	Value
hadoop.security.group.mapping	org.apache.hadoop.security.RuleBasedLdapGroupsMapping
hadoop.security.group.mapping.ldap.conversion.rule	to_lower

Note: This issue only occurs when Cloudera Base on premises is configured to convert names to lowercase, deviating from the default behavior that retains mixed case formatting. Reverting Cloudera Base on premises to its default configuration could resolve the issue but would require modifying Ranger policies, potentially causing downtime and significant effort for production environments. For customers unwilling to make such changes, the Cloudera Data Warehouse workaround is a practical solution.

DWX-21371: Incorrect end-of-support date in the Environments tab

In the Environments tab, a warning message may display an invalid end-of-support date: Mon Jan 01 0001 01:16:20 GMT+0116. This is a display-only issue and does not impact the actual end-of-support date for your environment or its functionality.

None.

DWX-21209: Resource template validation failure

When attempting to copy certain predefined resource templates, a validation error occurs, stating memory (should have a 33% overhead): X cannot be less than or equal to xmx*1.33: 30345. This issue affects the following templates:

- **Database Catalog**: Default resources, Medium resources, and Large resources
- **Impala**: Default resources
- **Hive**: Default resources

To resolve this, create a copy of the affected predefined templates, set them as default, and then apply the following changes:

- For Impala/Hive default templates, change the standalone query executor's Xmx to 88G.
- For Database Catalog default template, change Xmx to 6G or raise memory to 8555MB.
- For Database Catalog medium template, change Xmx to 12G.
- For Database Catalog large template, change Xmx to 21056M.

DWX-21185: Zookeeper configuration validation error during Cloudera Data Warehouse 1.5.5 activation

If TLS communication is not enabled in the Zookeeper service in Cloudera Base on premises, then in Cloudera Data Warehouse on premises version 1.5.5, the environment activation fails with the following Zookeeper configuration validation error:

```
unable to install dbcatalog charts: ssl.clientAuth not specified
 in zoo.cfg, so mutual tls authentication required for zookeeper
 server, but it is not supported in Cloudera Data Warehouse
```

You have to add ssl.clientAuth=none to the zookeeper_config_safety_valve in the Zookeeper service configuration in Cloudera Base on premises, and then retry the environment activation in Cloudera Data Warehouse on premises.

Note: Cloudera recommends enabling TLS communication in the Zookeeper server to secure communication between the Zookeeper server and Cloudera Data Warehouse components.

Known issues identified in 1.5.4

DWX-19477: Pods are stuck in pending state when you activate an environment with quota management enabled

Cloudera Data Visualization gets stuck in the pending state and wait for allocation when you activate an environment with quota management enabled because of a bug in the resource calculation for the Cloudera Data Visualization instance that is created from the Cloudera Data Warehouse UI. You may observe the following output when you run the `kubectl get pods` command:

```
kubectl get pods -n viz-rand-uru
NAME                                 READY   STATUS      RESTARTS
   AGE
service-discovery-56cc8ddc94-jpr5m   1/1     Running     0
   6m4s
viz-webapp-0                         0/1     Pending     0
    5m46s
viz-webapp-vizdb-create-job-588bs    0/1     Completed   0
   6m3s
```

Disable quota management before creating a Cloudera Data Visualization instance from the Cloudera Data Warehouse service.

DWX-18558: The executor pods in Impala Virtual Warehouse do not update when you change it to a different resosurce template

Suppose you created an Impala Virtual Warehouse with a certain resource template. If you apply a different resource template later having a different local storage size, the operation fails silently and the following pods are not updated: hiveserver 2, impala-coordinator, impala-executor and hue-backend. This happens because changing the storage size is not supported by Kubernetes.

None. Cloudera recommends that you avoid changing resource templates with different volume sizes and select the right size while creating the Virtual Warehouse.

Known issues identified in 1.5.3

DWX-17880: Hive Virtual Warehouse does not start if the bind user contains special characters

The Hive virtual warehouse may fail to start up if you have specified the following special characters in the LDAP bind credential password: < > & ' ". This happens because the HiveServer2 (HS2) pod gets stuck in the CrashLoopBackOff state with the following error in its logs: error parsing conf file:/etc/hive/conf/hive-site.xml com.ctc.wstx.exc.WstxUnexpectedCharException: Unexpected character '&' (code 38) in content after '<' (malformed start element?). at [row,col,system-id]: [388,13,"file:/etc/hive/conf/hive-site.xml"].

1. Change the LDAP bind credentials in the Cloudera Management Console. ensure that they do not contain the following unsupported special characters: < > & ' ".
2. Reactivate the environment in Cloudera Data Warehouse.

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 Cloudera Data Warehouse:

```
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-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 on premises.

Known issues identified in 1.5.5

DWX-20916: Refresh, Rebuild, and Update are disabled on pre-upgrade DBC and Virtual Warehouses

The Refresh, Rebuild, and Update operations have been disabled on DBC and Virtual Warehouse (VW) instances created before the upgrade from Cloudera Data Warehouse on premises 1.5.4 SP1 and earlier versions. This change was made to prevent service failures that previously occurred when these operations were executed.

Such failures included container errors and startup issues in Hue Query Processor and other Hue-related pods, caused by permission and directory-related problems.

Upgrade the affected DBC or Virtual Warehouse to complete the runtime upgrade. This resolves compatibility issues and allows services to function as expected.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4 SP1

DWX-19327: Unable to upgrade the Cloudera Data Visualization instance

If you are upgrading your Cloudera Data Visualization instance from Cloudera Data Warehouseon premises 1.5.4 or earlier versions to Cloudera Data Warehouse on premises 1.5.4 SP1, the upgrade fails with a “Failed to acquire lease” error.

This issue occurs because of a missing vizCRN from some of the older Cloudera Data Visualization application objects.

Create a new Cloudera Data Visualization instance. For more information, see Deploying Cloudera Data Visualization instance in Cloudera Data Warehouse.

Note: If you do not want to create a new instance, contact Cloudera support for further assistance.

Known issues identified in 1.5.4

DWX-18447: Virtual Warehouses go into an erroneous state after upgrading the Cloudera Control Plane

After upgrading the Cloudera Control Plane to Cloudera Data Services on premises 1.5.4, you may see that the Virtual Warehouse creation fails with the following error:

```
huefrontend-5888cc97b8-fxpx5 pod in compute-1716270252-fzfj
 namespace has an erroneous container, state: waiting, reason:
 RunContainerError, message: context deadline exceeded Error
 Code : undefined
```

When you run the `kubectl describe` command on the pod, you see the following warning:

```
Error: failed to create containerd task: failed to create shim
 task: OCI runtime create failed: runc create failed: unable to
 create new parent process: namespace path: lstat /proc/0/ns/ipc:
 no such file or directory: unknown
```

None. The Virtual Warehouse transitions back into the running or stopped state after the pod has initialized successfully.

Known issues on OpenShift cluster environments

This topic describes the OpenShift cluster environment known issues for Cloudera Data Warehouse on premises.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4

DWX-18151: Hue backup and restore fails on OpenShift Container Platform (OCP)

The file permissions for the pgpass file are set incorrectly on mount, due to restricted Security Context Constraint on OCP. This causes failure in connecting to the Hue database, resulting in failure of the job.

None.

ECS cluster environments

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

Known issues identified in 1.5.5

DWX-21037: Invalid trust store provided by Cloudera Data Warehouse when certification manager is enabled

When the Certification Manager is enabled, the trust store file provided by the Cloudera Data Warehouse becomes invalid for connecting to Beeline.

You can fetch a trust store file from the Cloudera Data Warehouse UI to configure TLS and connect to Beeline. For instructions on downloading the trust store file, see Downloading root certificates from Cloudera Data Warehouse web UI.

However, when the certification manager is enabled, Cloudera Data Warehouse continues to provide the same trust store file, even though the Hive cluster uses a different certificate. The issue arises because Cloudera Data Warehouse cannot provide a trust store compatible with the configured certificate issuer.

To resolve this issue, use your own certificate authority and set up a trust store on your system. You can do this by fetching the public key of the root CA (in .pem or .crt format), and then creating a new trust store using the following command:

```
keytool -importcert -trustcacerts -file rootca.pem -alias rootca
 -keystore truststore.jks -storepass changeit
```

Known issues identified in 1.5.4

No new known issues identified in 1.5.4.

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 on premises.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4

COMPX-18140: Database Catalog does not get created after activating the Cloudera Data Warehouse Environment

When you deactivate and reactivate a Cloudera Data Warehouse Environment on which you are using deterministic namespaces, you may notice the “namespace already exists. error DB Catalog already exists (cause: namespace resource exists with name <warehouse>, Kubernetes resource not unique error)” error and the Database Catalog is not created automatically.

Wait for a few minutes after deactivating and before reactivating the Cloudera Data Warehouse Environment.

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 Cloudera Data Warehouse 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 on premises.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4

DWX-18505: Resource template changes back to Reduced resources while creating a Virtual Warehouse

On the **Create Virtual Warehouse** modal, suppose you selected Default resources from the Resource Template drop-down menu. When you wait for a few seconds or click elsewhere on the modal, you see that the resource template changes back to “Reduced resources”. However, when you create the Virtual Warehouse, the Virtual Warehouse is created with the resource template you initially selected. This is a Cloudera Data Warehouse UI defect you encounter on environments that are activated using the low resource mode.

None.

DWX-18445: Virtual Warehouses get stuck in the deleting state

You may intermittently notice that a Virtual Warehouses gets stuck in the “deleting” state when you delete a Virtual Warehouse in Cloudera Data Warehouse. This could be due to a failing diagnostic job that is in an “Init:0/1” state in the namespace.

None. The diagnostic job gets cleaned up in one hour, after which the Virtual Warehouse gets deleted successfully.

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 Cloudera Base on premises 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 Cloudera 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-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 Cloudera Data Warehouse 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-site.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 Cloudera Data Warehouse 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 in Impala Virtual Warehouses

This topic describes the Impala Virtual Warehouse known issues for Cloudera Data Warehouse on premises.

Known issues identified in 1.5.5

DWX-21021: Impala virtual warehouses does not function after host reboot

After a host restarts, Impala components struggle to resume operations on the same node if the cluster has limited resources. This happens because these components are tied to specific local storage. If other small pods occupy that storage first, Impala's main processing units get stuck and cannot start.

Rebuild the affected virtual warehouses. See, Rebuilding a Virtual Warehouse

DWX-21147: Virtual Warehouse Creation Fails Intermittently After Upgrade

After upgrading from version Cloudera Data Warehouse on premises 1.5.4 CHF3 to 1.5.5, the first attempt to create or update an Impala Virtual Warehouse might fail. This occurs due to an internal timing conflict during resource pool migration.

Retry the Virtual Warehouse creation or update.

Known issues identified in 1.5.4

DWX-18505: Resource template changes back to Reduced resources while creating a Virtual Warehouse

On the **Create Virtual Warehouse** modal, suppose you selected Default resources from the Resource Template drop-down menu. When you wait for a few seconds or click elsewhere on the modal, you see that the resource template changes back to “Reduced resources”. However, when you create the Virtual Warehouse, the Virtual Warehouse is created with the resource template you initially selected. This is a Cloudera Data Warehouse UI defect you encounter on environments that are activated using the low resource mode.

None.

DWX-18445: Virtual Warehouses get stuck in the deleting state

You may intermittently notice that a Virtual Warehouses gets stuck in the “deleting” state when you delete a Virtual Warehouse in Cloudera Data Warehouse. This could be due to a failing diagnostic job that is in an “Init:0/1” state in the namespace.

None. The diagnostic job gets cleaned up in one hour, after which the Virtual Warehouse gets deleted successfully.

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 Cloudera Data Warehouse 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.

Known issues in Hue

This topic describes the Hue known issues for Cloudera Data Warehouse on premises.

Known issues identified in 1.5.4 SP1

DWX-19016: Hue Importer displays an incorrect status message

When you create an Impala table by importing CSV files using the Hue Importer, the Importer window might display a warning indicating that the query has failed. This is an incorrect message and the table is successfully created.

None. This issue is only related to the display of an incorrect status message and does not affect the actual table creation process.

Known issues identified in 1.5.4

CDPD-69394: Hue does not display logs while creating a table using the Importer

When you import a file in Hue using the Importer to create a table, the create table query is triggered and the operation succeeds. However, you may intermittently notice Hue not displaying the logs on the web interface.

None. This issue does not impact the underlying operation.

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 Cloudera Data Warehouse on premises.

Known issues identified in 1.5.0

DWX-12616: Hue limitation in Cloudera Data Warehouse on premises

Following are the known limitations in Hue in Cloudera Data Warehouse on premises 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 Cloudera Data Warehouse on premises if HDFS is configured for High Availability on the base cluster. Currently, Hue in Cloudera Data Warehouse 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 Cloudera Data Warehouse 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 Cloudera 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 Cloudera 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 Cloudera 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 on premises.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4

DWX-18475: Changing the resource template does not update the query executor and coordinator pods

If you have enabled ETL Isolation on an Unified Analytics Virtual Warehouse, changing the resoure template does not update the query executor and coordinator pods. They retain the initial values from the resource template that was used when you created the Virtual Warehouse.

Create a new Virtual Warehouse with the desired resource template.

DWX-18143: Unified Analytics-specific pods do not get updated after changing the resource template

When you change the resource template of an Impala Virtual Warehouse with the Unified Analytics option enabled from the **Virtual Warehouses Details** page, Cloudera Data Warehouse only updates the resources for Impala coordinator, statestore, catalogd pods, and so on. It does not update the resources for the HiveServer2 (HS2), Hive query executor, Hive coordinator, and standalone query executor pods.

None.

Iceberg-related known issues in Cloudera Data Warehouse on premises

This topic describes the Iceberg-related known issues in Cloudera Data Warehouse on premises.

Known issues identified in 1.5.5

No new known issues identified in 1.5.5

Known issues identified in 1.5.4 SP1

Hive compaction of Iceberg tables results in a failure

When Cloudera Data Warehouse and Cloudera Base on premises are deployed in the same environment and use the same Hive Metastore (HMS) instance, the Cloudera Base on premises compaction workers can inadvertently pick up Iceberg compaction tasks. Since Iceberg compaction is not yet supported in the latest Cloudera Base on premises version, the compaction tasks will fail when they are processed by the Cloudera compaction workers.

In such a scenario where both Cloudera Data Warehouse and Cloudera Base on premises share the same HMS instance and there is a requirement to run both Hive ACID and Iceberg compaction jobs, it is recommended that you use the Cloudera Data Warehouse environment for these jobs. If you want to run only Hive ACID compaction tasks, you can choose to use either the Cloudera Data Warehouse or Cloudera Base on premises environments.

If you want to run the compaction jobs without changing the environment, it is recommended that you use Cloudera Data Warehouse. To avoid interference from Cloudera Base on premises, change the value of the hive.compactor.worker.threads Hive Server (HS2) property to '0'. This ensures that the compaction jobs are not processed by Cloudera Base on premises.

1. In Cloudera Manager, click Clusters Hive Configuration to navigate to the configuration page for HMS.
2. Search for hive.compactor.worker.threads and modify the value to '0'.
3. Save the changes and restart the Hive service.

DWX-19489: Concurrent Hive-Iceberg UPDATE/INSERT query fails

Concurrent UPDATE/INSERT queries on Hive Virtual Warehouses might fail intermittently with the following error:

```
return code 40000 from org.apache.hadoop.hive.ql.exec.MoveTask.
Error committing job
```

Run the failed queries again.

Known issues identified in 1.5.4

No new known issues identified in 1.5.4.

Known issues identified in 1.5.2

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 Cloudera Data Warehouse data service associated with a Cloudera Base on premises 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 Cloudera Base on premises 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 Cloudera Data Warehouse nor does it affect creating of policies in Ranger.

On Cloudera Base on premises 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 Cloudera Base on premises 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:

```
{
  "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.

Technical Service Bulletins

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in Cloudera Data Warehouse

Cloudera Data Warehouse customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in Cloudera Data Warehouse.

Fixed issues in Cloudera Data Warehouse on premises

Review the issues fixed in this release of the Cloudera Data Warehouse service.

DWX-20188: Impala Virtual Warehouse with Unified Analytics mode triggers 'KubeStatefulSetReplicasMismatch' errors

When creating an Impala Virtual Warehouse within Cloudera Data Warehouse without setting the Active-Active HA mode, error messages are displayed by the Monitoring application, either on the OpenShift Alarm page or in the Cloudera Management Console Dashboard with the 'KubeStatefulSetReplicasMismatch' message.

This issue has been resolved, and the 'KubeStatefulSetReplicasMismatch' errors no longer appear.

DWX-20849: Multiple stateful set pods failing failing after node restart

After a node restart, some Cloudera Data Warehouse stateful set pods may become stuck in the initializing phase and display the following error:

```
Unable to retrieve some image pull secrets (docker-image-pull-se
cret); attempting to pull the image may not succeed.
```

The image pull requests are removed in the current release and this issue is no longer observed.

Preview features in Cloudera Data Warehouse on premises

This release of the Cloudera Data Warehouse service on premises introduces this technical preview.

Note: Technical previews are considered under development. Do not use these features in production environments.

Running queries on system tables (Preview)

Queries against Impala system tables, such as sys.impala_query_live, could get delayed due to admission control constraints. These queries, which require only coordinator resources, were previously blocked by queries competing for executor resources. To address this, Impala introduces an "only coordinators" request pool, allowing system table queries to bypass executor queues and run only on the coordinators to prevent delays during admission. This feature is Technical Preview except for workload aware autoscaling virtual warehouses where it is not supported. For more information, see Running queries on system tables

Apache Jira: IMPALA-13201

User quotas in admission control (Preview)

This release introduces user quotas in Impala admission control, a new feature designed to enhance resource management and ensure fair query distribution across users and groups. This feature is currently in Technical Preview and is not supported for virtual warehouses with workload-aware autoscaling.

For more information, seeUser quotas in Admission Control

Behavior changes in Cloudera Data Warehouse on premises

This release of the Cloudera Data Warehouse service on premises has the following behavior changes:

Summary: Default value for fe_service_threads increased to improve concurrency

Before this release: The default value for the fe_service_threads setting was 96.

After this release: Starting with Cloudera Data Warehouse on on premises 1.5.5, the default value is 128.

Summary: Simplified Cloudera Data Warehouse Diagnostic Bundle Download Process

The diagnostic bundle download process in Cloudera Data Warehouse has been simplified for an improved user experience.

Before this release: Users had to select specific information types within time intervals or choose a custom time interval. Additionally, they needed to manually adjust options in "Collect For" to include or exclude types of logs for the bundle.

After this release: Users now directly access a simplified "Collect" option, eliminating the need for additional time interval and log selection adjustments.

Version information for Cloudera Data Warehouse on premises components

Cloudera Data Warehouse uses Hive, Impala, and Hue as its Runtime components and also provides integration with Cloudera Data Visualization. Review the version information of Cloudera Data Warehouse on premises 1.5.5 components.

Table 1: Cloudera Data Warehouse on premises version information

Cloudera Data Warehouse component	Version
Hive	3.1.3000.2025.0.19.1-74
Impala	4.4.0.2025.0.19.1-74
Hue	4.5.0.2025.0.19.1-74
Cloudera Data Visualization	7.2.7-b48
Cloudera Data Warehouse server	1.11.0-b190
CDP CLI	0.9.139

Apache Iceberg version information

The following table shows the version of the Iceberg component in this release of Cloudera Data Warehouse on premises:

Cloudera Data Warehouse server version	Cloudera Data Warehouse Runtime version	Iceberg version
1.11.0-b190	2025.0.19.1-74	1.5.2.2025.0.19.1-74

Cumulative hotfixes: Cloudera Data Warehouse

Review the cumulative hotfixes that have been shipped for Cloudera Data Warehouse on premises 1.5.4.

Cloudera Data Warehouse 1.5.5-CHF1

There were no features, fixes, and known issues in the Cloudera Data Warehouse 1.5.5 Cumulative hotfix 1 release.

Service pack releases

Review the list of service pack releases that were shipped for Cloudera Data Warehouse on premises.

Cloudera Data Warehouse on premises 1.5.5 SP1

Version information for Cloudera Data Warehouse on premises 1.5.5 SP1 components

Cloudera Data Warehouse uses Hive, Impala, Trino, Hue as its Runtime components and also provides integration with Cloudera Data Visualization. Review the version information of Cloudera Data Warehouse on premises 1.5.5 SP1 components.

Note: Trino is in Technical Preview and is not ready for production deployments. Cloudera recommends trying this feature in test or development environments and encourages you to provide feedback on your experiences.

Table 2: Cloudera Data Warehouse on premises version information

Cloudera Data Warehouse component		Version
Cloudera Data Warehouse server		1.11.0-h1000-b133
Cloudera Data Visualization		8.0.5.1000-b4
CDP CLI		0.9.150
Cloudera Data Warehouse Runtime		2025.0.19.1000-43
Cloudera Data Warehouse Components	Hue	4.5.0.2025.0.19.1000-43
	Hive	3.1.3000.2025.0.19.1000-43
	Iceberg	1.5.2.2025.0.19.1000-43
	Impala	4.4.0.2025.0.19.1000-43
	Trino	0.451.1.2025.0.19.1000-43

What's new in Cloudera Data Warehouse on premises 1.5.5 SP1

Review the new features introduced in this cumulative hotfix release of Cloudera Data Warehouse on premises 1.5.5 SP1.

- Cloudera Data Warehouse
- Hive
- Impala
- Trino
- Hue

Cloudera Data Warehouse on premises

Deactivating environments

The Cloudera Data Warehouse user interface now includes a Force Delete option for environments that are stuck in a deleting state. This feature empowers administrators to directly remove such environments from the UI, bypassing the standard deletion process. For more information on removing the environments, refer to Deactivation environment.

Auto-scaling enhancements

In this release, the auto-scaling behavior for Hive and Impala Virtual Warehouses is enhanced to ensure more reliable and efficient resource allocation. With this update, the system now proactively reserves the requested quota, the resources needed for the maximum number of executor groups configured. This guarantees that during a scale-out event, new executor groups are created and Yunikorn schedules them based on the availability of resources at that moment. For more information, refer to Quota management in Cloudera Data Warehouse on premises.

Rebuild-restore method and its limitations

The rebuild-restore method, which uses the cluster backup-restore functionality, has some limitations. To successfully restore a selected namespace, you must first delete Cloudera Data Visualizations, Impala, and Hive Virtual Warehouses. This requirement does not apply to Database Catalogs. For more information, refer to Using DRS with Cloudera Data Warehouse.

Manual resource pool assignment at upgrade

In this release, you can now manually assign resource pools to existing Virtual Warehouses, and Cloudera Data Visualizations after upgrading to 1.5.5 SP1. This new functionality is for entities that were not yet enabled for quota management. For more information, refer to Adding resource pools after Cloudera Data Warehouse 1.5.5 SP1 upgrade.

Optimized node allocation for query executor pods

In this release, you can enhance performance and resource utilization by dedicating nodes with local storage exclusively for Cloudera Data Warehouse query executor pods. This feature is disabled by default. When enabled, Hive and Impala executor and coordinator pods are scheduled exclusively on these specifically labeled worker nodes, ensuring that other data services do not utilize them.

Cloudera Data Visualization upgrade feasibility and benefits

Cloudera Data Visualization has been upgraded to version 8.x, offering enhanced security, such as addressing numerous CVEs and greater reliability due to the migration to Chainguard. This new version also introduces advanced capabilities, such as AI Visual features. For more details, refer to the Cloudera Data Visualization guide.

What's new in Hive on Cloudera Data Warehouse on premises

Common table expression detection and rewrites using cost-based optimizer

Hive's existing shared work optimizer detects and optimizes common table expressions heuristically, but it lacks cost-based analysis and has limited customization. Introduced new APIs and configuration options to support common table expression optimizations at the cost-based optimizer level. The feature is experimental and disabled by default.

Apache Jira: HIVE-28259

Upgraded Avro to version 1.11.3

What's new in Impala on Cloudera Data Warehouse on premises

OpenTelemetry integration for Impala

Impala now has OpenTelemetry (OTel) support to help you see query performance and troubleshoot issues. This new feature, available in Cloudera Data Warehouse on premises 1.5.5 SP1, collects and exports query telemetry data as OpenTelemetry traces to a central OpenTelemetry compatible collector.

To enable this, you must upgrade your existing Impala Virtual Warehouses after the Cloudera Data Warehouse version upgrade. The integration is designed to have a minimal impact on performance because it uses data already being collected and handles the export in a separate process. For more information, see OpenTelemetry support for Impala.

Apache Jira:IMPALA-13234

Enable global admission controller

A new single admission controller service has been added to Impala to improve performance in multi-coordinator setups. It is now a separate service, which prevents its failure from affecting coordinators and executors. This feature is enabled by default for new Impala Virtual Warehouses running in High Availability (HA) Active-Active mode. If needed, you can disable it through the Cloudera web interface, but this action is permanent. For more information, see Impala admissiond and Configuring admission control.

What's new in Trino on Cloudera Data Warehouse on premises

Introducing support for Trino Virtual Warehouses [Technical Preview]

Cloudera Data Warehouse now supports the creation and management of Trino Virtual Warehouses. For information about creating a Trino Virtual Warehouse, see Adding a new Virtual Warehouse.

Trino is a distributed SQL query engine designed to efficiently query large datasets across one or more heterogeneous data sources. This integration enables users to leverage Trino's powerful capabilities directly within Cloudera Data Warehouse.

With this integration, you can configure and deploy Trino connectors to seamlessly connect to diverse remote data sources, access data, expose metadata, and manage data transfer to and from the remote sources. For more information, see Trino Federation Connectors.

Authorization for Trino is supported through Apache Ranger by default through the cm_trino authorization service. You can create or update Ranger policies for specific resources and assign permissions to Trino users, groups, or roles. When a user submits a query to Trino, the system verifies the defined policies to ensure that the user has the necessary permissions to run queries. For more information, see Ranger authorization for Trino Virtual Warehouses.

Note: Trino is in Technical Preview and is not ready for production deployments. Cloudera recommends trying this feature in test or development environments and encourages you to provide feedback on your experiences.

What's new in Hue on Cloudera Data Warehouse on premises

General availability of deploying a shared Hue service

Cloudera Data Warehouse now supports the deployment of a shared Hue service, enabling cost-efficient management by ensuring that only the necessary Virtual Warehouses remain active. Organisations can enhance team isolation by running multiple shared Hue instances, providing flexibility and control. The shared Hue service remains available as long as the environment is active.

For more information, see About deploying the shared Hue service.

New Hue storage browser (Technical Preview)

The Hue Storage Browser is a web-based interface designed to provide seamless interaction with multiple file systems. With enhanced usability and functionality, the File Browser improves data management, offering a streamlined experience.

For more information, see Hue Storage Browser.

Enhanced file extension controls for Hue file upload

Earlier, Hue permitted uploading all file types to the configured filesystems, including unsupported extensions, which posed a security risk.

To enhance security, Hue now allows you to control both allowed and restricted file type extensions for Hue file uploads. This improves flexibility and enhances the overall user experience.

For more information, see Managing file extensions for Hue uploads.

Fixed issues in Cloudera Data Warehouse on premises 1.5.5 SP1

Review the issues fixed in this release of the Cloudera Data Warehouse service.

DWX-21395: Log routers are not requesting enough resources

Previously, the Cloudera Data Warehouse log-router component's resource quota was insufficient. Because the log router runs as a DaemonSet (with instances on multiple nodes), the resource request calculation did not account for the number of nodes or pod instances in the cluster. This led to resource constraints and issues with log router pods.

This issue is now resolved ensuring that sufficient resources are requested and allocated. Cloudera Data Warehouse on premises now correctly calculates the resource quota of the log router by multiplying the resource request of the underlying pod by the number of nodes it is executed on.

DWX-14325: Support for virtual warehouse access control in Cloudera Data Warehouse on premise for Hive with Kerberos

Previously, enabling warehouse-level access control for Hive or Impala warehouses in Unified Analytics mode and associating a user group with a Virtual Warehouse disabled Kerberos

authentication. This prevented the use of Kerberos security alongside warehouse-level access control, limiting the ability to enforce both access restrictions and strong authentication simultaneously.

This issue is now resolved. When warehouse-level access control is enabled for Hive Virtual Warehouses, Kerberos authentication remains functional. This ensures that only the designated user groups can access the Virtual Warehouse through supported channels such as Hue, JDBC, Beeline, Impala-shell, and Impyla, while maintaining the security provided by Kerberos authentication.

DWX-21185: Resolution of Zookeeper configuration validation error during Cloudera Data Warehouse 1.5.5 activation

When activating a Cloudera Data Warehouse 1.5.5 on premises, the process failed with an incorrect Zookeeper configuration validation error related to ssl.clientAuth. This error occurred specifically when TLS communication was not enabled for the Zookeeper service in the Cloudera Base on premises deployment.

This issue is now resolved, and activation now works regardless of the TLS configuration.

DWX-22116: Upgrade failures due to certificate error

Previously, upgrading from version 1.5.5 to 1.5.5 service pack release could fail due to an underlying certificate verification error. The system's internal components, such as the cdp-release-resource-pool-manager pod, encountered the error x509: certificate signed by unknown authority, which prevented successful communication and operations.

This issue is now resolved. The system is fixed to correctly handle and verify the necessary certificates during and after the upgrade process.

DWX-21037: Invalid trust store provided by Cloudera Data Warehouse when certification manager is enabled

Previously, when the certi-manager was enabled, the trust store file provided by the Cloudera Data Warehouse UI was invalid for Beeline connections due to a mismatch with the Hive cluster certificate.

This issue is now resolved, and the UI now provides a compatible trust store file, allowing users to configure TLS and connect to Beeline successfully.

DWX-21209: Resource template validation failure

Previously, when attempting to copy certain predefined resource templates, a validation error occurred with the memory (should have a 33% overhead): X cannot be less than or equal to xmx*1.33: 30345 message.

This issue is now resolved, and the validation error no longer occurs.

DWX-2147: Virtual warehouse creation fails intermittently after upgrade

Previously, after upgrading Cloudera Data Warehouse on premises to version 1.5.5, the first attempt to create or update an Impala Virtual Warehouse might have failed due to an internal timing conflict during resource pool migration.

This issue is now resolved.

DWX-15302: Upgrade button stays visible even after the upgrade completes

Previously, after upgrading the Database Catalog, the Upgrade button remained visible on the Cloudera Data Warehouse web interface instead of disappearing or being disabled.

This issue is now resolved.

CDPD-83530: Task commits were allowed despite an exception being thrown in the Tez processor

A communication failure between the coordinator and executor caused a running task to terminate, resulting in a java.lang.InterruptedException being thrown by the ReduceRecordProcessor.init(). Despite this exception, the process still allowed the task to be committed and generated a commit manifest.

This issue has now been resolved. The fix ensures that outputs are not committed if an exception is thrown in the Tez processor.

Apache Jira: HIVE-28962

Cookie-Based authentication support for JWT tokens

When JWT tokens are used for authentication, every HTTP request within a session requires token verification. If these tokens have a short lifespan, it can lead to authentication failures and disrupt session continuity.

This issue is now resolved by using authentication cookies, which generally have a longer lifespan (configured through the max_cookie_lifetime_s flagfile option) and can remain valid for the duration of the session. This enables subsequent authentication requests to rely on cookies rather than repeatedly verifying the JWT token.

Apache Jira: IMPALA-13813

CDPD-80798: Stable Catalogd initialization in HA mode

Catalogd initialization previously might timeout to complete in high availability mode. This happened because metadata operations started prematurely, blocking Catalogd from becoming active.

This issue is resolved by ensuring Catalogd determines HA state before starting metadata operations in HA mode. This prevents blocking issues and ensures a stable startup.

Apache Jira: IMPALA-13850

CDPD-83059: Optimized Impala Catalog cache warmup

Impala's Catalogd previously started with an empty cache. This led to slow query startup for important tables and affected high availability failovers.

This issue is resolved by adding new settings to pre-load specific tables into the Catalogd cache in the background. This ensures faster query startup and smoother high availability failovers.

Apache Jira: IMPALA-14074

CDPD-87222: Consistent TRUNCATE operations for external tables

Impala's TRUNCATE operations on external tables previously did not consistently delete files in subdirectories, even when recursive listing was enabled.

This issue is resolved by ensuring Impala uses the HMS API for TRUNCATE operations by default.

Apache Impala: IMPALA-14189, IMPALA-14224

CDPD-82415: TABLESAMPLE clause of the COMPUTE STATS statement has no effect on Iceberg tables

This fix resolves a regression introduced by IMPALA-13737. For example, the following query scans the entire Iceberg table to calculate statistics, whereas it should ideally use only about 10% of the data.

```
COMPUTE STATS t TABLESAMPLE SYSTEM system(10);
```

This fix introduces proper table sampling logic for Iceberg tables, which can be utilized for COMP UTE STATS. The sampling algorithm previously located in IcebergScanNode.getFilesSample() is now relocated to FeIcebergTable.Utils.getFilesSample().

Apache Jira: IMPALA-14014

CDPD-82599: Query rejected due to too many fragment instances on one node

Some queries failed with a scheduling error when too many fragment instances were placed on a single executor node.

The fix limits the number of instances per node during scheduling to avoid exceeding the maximum allowed.

Apache Jira: IMPALA-14006

CDPD-80939: Missing equivalence conjunct in aggregation node with inline views

In queries that include filters on the result of a UNION operation, the planner sometime removed required conjuncts, which caused incorrect query results.

This has now been fixed.

Note: This issue affected only distinct UNION operations, not UNION ALL.

Apache Jira: IMPALA-13873

CDPD-82303: EXEC_TIME_LIMIT_S incorrectly includes planning time

The EXEC_TIME_LIMIT_S timer was triggered during the planning and admission(queuing) phases, which could cause queries to fail before any processing on backend started.

The issue was addressed by starting the EXEC_TIME_LIMIT_S only after the query is ready to run on the backend. This ensures the timer applies only to the actual processing phase.

Apache Jira: IMPALA-14001

CDPD-82364: Nested loop join rewrites disjunctive subquery incorrectly

Queries with a subquery inside an OR condition could return incorrect results due to an improper join predicate rewrite.

The issue was addressed by skipping the join rewrite when the subquery is part of a disjunctive (OR) expression.

Apache Jira: IMPALA-13991

Invalid cardinality calculation in sortnode's computestats

An error occurred during query execution due to an overflow in the calculation of row limits, causing unexpected failures.

The calculation was updated to prevent overflow and eliminating the error.

Apache Jira: IMPALA-14070

TABLESAMPLE clause of the COMPUTE STATS statement has no effect on Iceberg tables

This fix resolves a regression introduced by IMPALA-13737. For example, the following query scans the entire Iceberg table to calculate statistics, whereas it should ideally use only about 10% of the data.

```
COMPUTE STATS t TABLESAMPLE SYSTEM system(10);
```

This fix introduces proper table sampling logic for Iceberg tables, which can be utilized for COMP UTE STATS. The sampling algorithm previously located in IcebergScanNode.getFilesSample() is now relocated to FeIcebergTable.Utils.getFilesSample().

Apache Jira: IMPALA-14014

IllegalStateException with Iceberg table with DELETE

Running a query on an Iceberg table fails with an IllegalStateException error in the following scenario:

- The Iceberg table has delete files for every data file (no data files without delete files) AND
- An anti-join operation is performed on the result of the Iceberg delete operation (IcebergDeleteNode or HashJoinNode)

This fix resolves the issue by setting the TableRefIds of the node corresponding to the Iceberg delete operation (IcebergDeleteNode or HashJoinNode) to only the table reference associated with the data files, excluding the delete files.

Apache Jira: IMPALA-14154

Error unnesting arrays in Iceberg tables with DELETE files

The following error occurred when unnesting a nested array (a 2D array) from an Iceberg table. This issue was triggered specifically when the table contained delete files for some, but not all, of its data files.

```
Filtering an unnested collection that comes from a UNION [ALL] is
 not supported yet.
```

Reading an Iceberg table with this mixed data and delete file configuration creates a UNION ALL node in the query execution plan. The system had a check that explicitly blocked any filtering on an unnested array.

This fix relaxes the validation check, allowing the operation to proceed if all UNION operands share the same tuple IDs. This ensures the query can successfully unnest the array.

Apache Jira: IMPALA-14185

DWX-21190: Hue UI can't upload files larger than 10Mb to Ozone

Previously, Hue UI experienced file upload failures for files larger than 10MB when uploading to Ozone. This was traced to incompatible chunked upload handling caused by constraints in Ozone's HttpFS API. Additionally, you needed to navigate to the Groups tab, select the default group, and manually enable the file browser.ofs_access: Access to OFS from filebrowser and filepicker perm ission to access the file browser.

The issue is now resolved by enabling Ozone auto-configuration by default, disabling the chunked uploader, and tuning the chunk size and timeout parameters accordingly.

CDPD-82819: Improved kt_renewer logging for kerberos ticket renewal failures

Previously, the kt_renewer application stopped after three retry attempts and did not log the kinit error, making troubleshooting difficult. This occurred because the subprocess output streams (stdout and stderr) were read twice, causing empty error messages to be recorded when kinit failed.

This issue is now resolved by enhancing the logging mechanism, which is improved to read and decode the output streams only once, ensuring that detailed kinit error messages are properly captured in the logs.

Known issues in Cloudera Data Warehouse on premises 1.5.5 SP1

Review the known issues and limitations that you might run into while using the Cloudera Data Warehouse service in Cloudera Private Cloud Data Services.

Known issues identified in 1.5.5 SP1

DWX-21588 - CLI timeout issues for the create-backup command

Due to network latency or cluster slowness, the create-backup command might exceed the default 60-second CLI timeout. This causes the CLI to retry the command, which might fail because of an existing Hue or Cloudera Data Visualization backup job on the cluster with the same name.

To prevent the timeout issue, set the --cli-read-timeout option to 0 in your command. This will disable the CLI timeout.

However, as the API has a hard limit of 200 seconds, the command will still return a timeout error in response if the hard limit is exceeded. Despite the error, the backup will run and complete successfully in the background. You can get the Backup CRN from the cdp dw list-backups command response.

DWX-20754: Invalid column reference in lateral view queries

The virtual column BLOCK__OFFSET__INSIDE__FILEfails to be correctly referenced in queries using lateral views, resulting in the error:

```
FAILED: SemanticException Line 0:-1 Invalid column reference 'BL
OCK__OFFSET__INSIDE__FILE'
```

To resolve this issue, you can either:

1. Set the configuration property hive.cbo.fallback.strategy to CONSERVATIVE for the specific query containing such lateral views.
2. Specify column names explicitly in the 'SELECT' statement instead of using SELECT * in the subquery involving the lateral view, with the NEVER fallback strategy.

DWX-20491: Impala queries fail with "EOFException: End of file reached before reading fully"

Impala queries fail with an EOFException when reading from an HDFS file stored in an S3A location. The error occurs when the file is removed. If the file is removed using SQL commands like `DROP PARTITION`, there may be a significant lag in Hive Metastore event processing. If removed by non-SQL operations, run `REFRESH` or `INVALIDATE METADATA` on the table to resolve the issue.

Run `REFRESH/INVALIDATE METADATA <table>;`

DWX-20490: Impala queries fail with “Caught exception The read operation timed out, type=<class 'socket.timeout'> in ExecuteStatement”

Queries in impala-shell fail with a socket timeout error in execute statement which submits the query to the coordinator. The error occurs when query execution takes longer to start mainly when query planning is slow due to frequent metadata changes.

Increase the socket timeout on the client side. Set --client_connect_timeout_ms to a higher value, e.g. add --client_connect_timeout_ms=600000 to the impala-shell command line.

CDPD-76644: information_schema.table_privileges metadata is unsupported

Querying the information_schema.table_privileges access control metadata for ranger is unsupported and a TrinoException is displayed indicating that the connector does not support table privileges.

None.

CDPD-76643/CDPD-76645: SET AUTHORIZATION SQL statement does not modify Ranger permissions

The following SQL statements do not dynamically modify the Ranger permissions:

```
CREATE SCHEMA test_createschema_authorization_user AUTHORIZATION
 user;
ALTER SCHEMA test_schema_authorization_user SET AUTHORIZATION use
r;
```

As an Administrator, you can authorize the permissions from the Ranger Admin UI.

CDPD-68246: Roles related operations are not authorized by Ranger Trino plugin

When you have row filter policy for same resource and same user in both cm_trino and cm_hive (Hadoop SQL) repos and the row filtering conditions are different, then on querying the table using that user returns empty response in the trino-cli.

Do not create row filter policies for the same resource and same user in different repos.

DWX-19626: Number of rows returned by Trino does not match with the Hive query results

If you are running an exact same query on both Hive and Trino engine that involves dividing integers, it was observed that the query results returned by Trino does not match with the query results returned by the Hive engine. This is due to a default behavior of Trino when dividing two integers. Trino does not cast the result into a FLOAT data type.

While performing a floating point division on two integers, cast one of the integers to a DOUBLE. For more information, see the Trino documentation.

Deprecation notice for Cloudera Data Warehouse on premises 1.5.5 SP1 components

Review the features and functionalities that have been or will be removed or deprecated in this release of Cloudera Data Warehouse on premises.

Unified Analytics is deprecated

The Unified Analytics feature in the Cloudera Data Warehouse on premises is now deprecated. It is no longer available through the user interface, but you can still create and manage Impala Unified Analytics virtual warehouses using the CDP CLI.

Behavior changes in Cloudera Data Warehouse on premises 1.5.5 SP1

Summary: Changes to Unified Analytics availability in Cloudera Data Warehouse on premise

Before this release: When creating a new Impala Virtual Warehouse, you could select the Enable Unified Analytics option.

After this release: The Enable Unified Analytics option is no longer available when creating a new Impala Virtual Warehouse. However, you can continue to create and manage Impala Unified Analytics virtual warehouses using the Cloudera CLI.

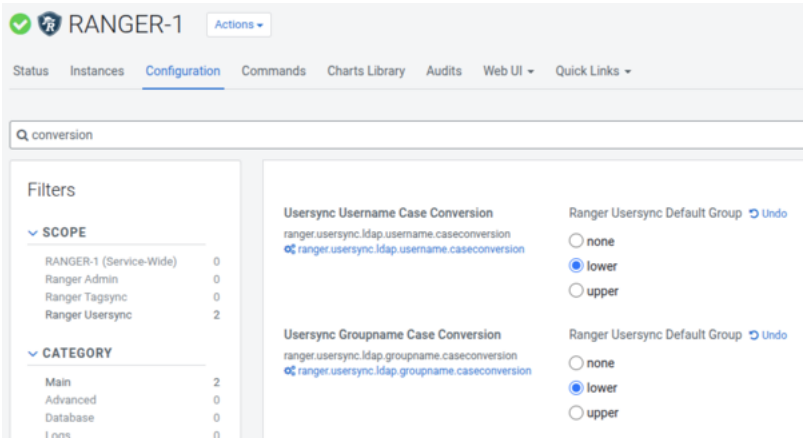
Summary: Enhanced log router resource allocation in Cloudera Data Warehouse on premise

Before this release: The resource quota of the log router component was fixed and static, failing to scale automatically with the cluster size. As clusters grew with additional nodes, the allocated quota became insufficient for the log router, which operates as a DaemonSet across all nodes. This resource deficit often caused the log router pods to remain in a Pending state and required manual quota adjustments to address the issue.

After this release: The resource quota calculation for the log router component is now improved. Cloudera Data Warehouse on premises now dynamically calculates and requests the appropriate quota by multiplying the resource request of a single log router pod by the current number of nodes it operates on. This ensures the log router consistently receives sufficient resources, eliminating the need for manual adjustments as the cluster scales.

Summary: Automatic synchronization of group name conversion between Cloudera Base on premises Ranger and Cloudera Data Warehouse

Before this release: In Active Directory environments, user and group names are often written in mixed case, for example, JohnDoe or AdminGroup, and handled in a case-insensitive manner by Windows. However, Cloudera Base on premises operates in a Linux environment, in which names are case-sensitive. To address this, some customers configure Cloudera Base on premises to disable case sensitivity in System Security Service Daemon (sssd) and modify Ranger Usersync settings to convert user and group names to lowercase, ensuring compatibility with Ranger policies.



While this configuration worked correctly in Cloudera Base on premises, authorization issues could arise in Cloudera Data Warehouse components like Hive and Impala. Cloudera Data Warehouse did not automatically convert group names to lowercase, causing mismatches with Ranger policies that define group names in lowercase. This resulted in authorization problems, such as users being unable to access databases, tables, or columns in Hue or remote client shells, such as impala-shell or jdbc, even though access worked correctly in Cloudera Base on premises Hue or remote client shells.

To resolve this issue, a manual workaround was required to enable group name conversion to lowercase in Cloudera Data Warehouse. This involved adding specific Hadoop core-site configuration entries to the hadoop-core-site-default-warehouse configuration file. For Hive Virtual Warehouse, changes were applied to HiveServer2 and for Impala Virtual Warehouse, changes were applied to Impala Catalogd, Impala Coordinator, Impala Executor, and Impala StateStored.

Property Name	Value
hadoop.security.group.mapping	org.apache.hadoop.security.RuleBasedLdapGroupsMapping
hadoop.security.group.mapping.ldap.conversion.rule	to_lower

After this release: The Cloudera Data Services on premises 1.5.5 SP1 release introduces automatic synchronization of group name conversion settings between Cloudera Base on premises and Cloudera Data Warehouse. This ensures consistent handling of group names across environments.

Behavior remains unchanged if the manual Cloudera Data Warehouse workaround was previously applied, or if the Cloudera Base on premises Ranger Usersync was set to the default of none, meaning no case conversion.

Action required: If mixed-case group name policies were created specifically for Cloudera Data Warehouse as an undocumented solution, these policies must be reviewed and manually updated to match the case conversion settings of being either lowercase or uppercase in Cloudera Base on premises Ranger.

Note: Failure to update mixed-case group name policies might result in mismatches and authorization issues after the release.

Summary: Support load-based routing in impala-proxy

Before this release: impala-proxy used a random selection policy to choose a coordinator to which to forward new OpenSession requests.

After this release: impala-proxy now uses load-based routing to distribute OpenSession requests across multiple active coordinators.

Summary: Cleanup subdirectories in truncate/insert overwrite if recursing listing is enabled

Before this release: Impala did not consistently delete files located in subdirectories of external tables during TRUN CATE and INSERT OVERWRITE operations, even when recursive listing was enabled. This led to leftover data in subdirectories after these operations, resulting in data corruption.

After this release:After this change, directories are also deleted in addition to (non-hidden) data files, with the exception of hidden and ignored directories. Now, setting DELETE_STATS_IN_TRUNCATE=false is no longer supported by default when truncating non-transactional tables; attempting this will result in an exception. If the old behavior is absolutely required, you can set the --truncate_external_tables_with_hms flag to false, but be aware that this will also reintroduce the bug that was fixed by this change.

Apache Impala: IMPALA-14189 IMPALA-14224