

Data Warehouse Release Notes

Date published: 2020-08-17

Date modified: 2022-11-18



Legal Notice

© Cloudera Inc. 2024. 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 Private Cloud?.....	4
November 18, 2022.....	4
September 16, 2022.....	6
June 21, 2022.....	6
April 13, 2022.....	8
January 14, 2022.....	8
October 29, 2021.....	9
October 4, 2021.....	9
April 28, 2021.....	10
December 16, 2020.....	14
August 17, 2020.....	16
 Known issues and limitations in Cloudera Data Warehouse Private	
Cloud.....	16
General known issues.....	16
Known issues on OpenShift cluster environments.....	17
ECS cluster environments.....	18
Known issues in Database Catalogs.....	18
Known issues in Hive Virtual Warehouses.....	18
Known issues in Impala Virtual Warehouses.....	19
Known issues in Hue.....	20
Data Analytics Studio.....	20
 Fixed issues in Cloudera Data Warehouse Private Cloud.....	21
 Runtime component versions for Cloudera Data Warehouse Private	
Cloud.....	21

What's new in Cloudera Data Warehouse Private Cloud?

This section lists major features and updates for the Cloudera Data Warehouse (CDW) service in CDP Private Cloud.

November 18, 2022

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services introduces the new features and improvements that are described in this section.

Ability to use deterministic namespace in Kerberos principals

CDW now uses a deterministic namespace and environment IDs. The Kerberos principals for Database Catalogs and Environments use the service hostname and the deterministic namespace name based on the name of the Database Catalog or Environment.

When you specify an Environment or Database Catalog name, CDW appends a prefix to the environment and Database Catalog name, as well as to the Kerberos principal name. For more information, see [Predefined Kerberos principals in Cloudera Data Warehouse Private Cloud](#).

Ability to install and manage Cloudera Data Warehouse clusters using CDP CLI

You can install CDP CLI on your computer and use it to install and manage CDW clusters on CDP Private Cloud. To install CDP CLI, see [CLI client setup](#). For the list of CDW sub-commands, see <https://cloudera.github.io/cdp-dev-docs/cli-docs/dw/index.html>.

Real-time graphs in Grafana display scratch and cache disk utilization

You can monitor and track the scratch and cache disk utilization at a cluster level and the Virtual Warehouse level. This enables you to size a Virtual Warehouse optimally and calculate the memory and disk requirements. For more information, see [Monitoring Data Warehouse service resources with Grafana dashboards](#).

Hue is now the unified next-generation SQL assistant in CDP

Hue packs the combined abilities of Data Analytics Studio (DAS) such as query optimization, query debugging framework, and rich query editor experience of Hue, making Hue the next-generation SQL assistant in CDP. You can search query history, view query details, visual explain plan, and DAG information, compare two queries, and download debug bundles for troubleshooting Hive queries from the **Job Browser** page. To support this feature, a new service called Query Processor is added to the CDP stack as a dependency for Hue.

A new tab called Hue query processor has been added under the Database Catalog Edit CONFIGURATIONS section. The hue-query-processor.json and hue-event-processor.json files are no longer available under the Virtual Warehouse Edit CONFIGURATIONS Hue Configuration files drop-down menu. For more information, see [About Hue Query Processor](#).

Data Analytics Studio (DAS) is deprecated

DAS is deprecated and is not installed by default. DAS will be unavailable in future releases. Cloudera encourages that you use Hue to run Hive LLAP workloads. If you need to use DAS, then you can enable it from the **Advanced Settings** page. See [Enabling Data Analytics Studio in CDW Private Cloud](#).

Ability to configure Impala coordinator high availability

You can configure up to five Impala coordinators in an active-active configuration concurrently with cookie-based load balancing to resolve or mitigate query concurrency problems. To enable the active-active configuration, select

the Enabled (Active-Active) option while creating a Virtual Warehouse. For more information, see [Configuring Impala coordinator high availability in CDW Private Cloud](#).

Ability to spill Impala queries to HDFS

You can configure heavy Impala queries to write intermediate files during large sorts, joins, aggregations, or analytic function operations to a remote scratch space on HDFS. To enable this feature, you must configure the Impala daemon to use the specified locations for writing the intermediate files and then specify the HDFS URI while creating the Impala Virtual Warehouse. For more information, see [Enabling Impala to spill to HDFS in CDW](#).

New Advanced Configurations menu for enabling and disabling deprecated and Technical Preview features

A new “Advanced Configurations” menu has been added to the CDW web interface which opens an **Advanced Settings** page. On this page, you can enable or disable Technical Preview and deprecated features which are not installed or available out of the box when you install the Private Cloud data services. For example, enabling DAS which has been deprecated and enabling third-party S3 providers in private cloud.

Using the Refresh option to apply configuration changes

A new Refresh option has been added to the more options menu for Database Catalogs and Virtual Warehouses that helps you to apply configuration changes that you made at an environment level, from the Management Console, or from the **Advanced Settings** page. In most cases, this helps you to avoid deleting and recreating Database Catalogs or Virtual Warehouses. To learn more about the use cases in which you can use the Refresh option, see [About the Refresh option](#).

Support added for AWS S3 and third-party object storage (Preview)

CDW supports using AWS S3 object storage services for storing tables. Other similar, compatible, on-premises object stores that support the S3 protocol can work as well. CDW exposes Hive and Impala tables stored on S3 as SQL tables which you can query using Hue. However, you cannot browse and import files to create tables from S3 in Hue. For more information, see [Third-party object storage support for CDW Private Cloud](#).

This feature is in Technical Preview and not recommended for production deployments. Cloudera recommends that you try this feature in test or development environments.

Ability to change delegation username and password

You can update the delegation username and password that CDW uses to impersonate authorization requests from Hue to the Impala engine from the **Environment Details** page. For more information, see [Changing delegation username and password](#).

Ability to configure Impala coordinator and executor pod size is GA

You can optimize the performance of your Impala Virtual Warehouse and resources used in an environment, based on your hardware configuration, by customizing the amount of resources allocated to the Impala coordinators, executors, and catalog daemons. This helps you to better leverage intra-query parallelism and achieve powerful compute clusters with fewer nodes.

This feature is generally available (GA) starting 1.4.1, and you can use it in production environments.

Earlier, CDW allowed you to specify the path and size for scratch and cache space for Impala executor and coordinator pods. Starting with CDW 1.4.1, you can only specify the size for these parameters.

A new parameter, “Overhead size” has been added which allows you to specify storage size for storing resources that are used by the tools run by the container. For more information, see [Creating custom pod configurations for Impala Virtual Warehouses](#).

Data Visualization integration in Cloudera Data Warehouse is GA

CDW integrates Data Visualization for building graphic representations of data, dashboards, and visual applications based on CDW data, or other data sources you connect to. Authorized users can explore data using graphics such as pie charts and histograms, and collaborate using dashboards. BI analysts who can access your environment can use these features. To get started with Cloudera Data Visualization, see [Creating a Data Visualization instance in CDW](#)

This feature is generally available (GA) starting 1.4.1, and you can use it in production environments.

SSL support for Oracle database is GA

CDW supports Oracle database and can connect to SSL-enabled Oracle on the base cluster. For optimum security, the network connection between the default Database Catalog Hive MetaStore (HMS) in CDW and the relational database hosting the base cluster's HMS must be encrypted with SSL. For more information, see [Configuring Oracle database to use SSL for Data Warehouse](#).

This feature is generally available (GA) starting 1.4.1, and you can use it in production environments.

Improved read performance of ORC tables by Impala

Continuous improvements in Impala's read performance of ORC tables.

Ozone filesystem support added for Hive and Impala (Preview)

You can use Apache Ozone storage with CDW Private Cloud. This feature is in Technical Preview and Cloudera recommends that you try this in test or development environments. For more information, see [Using Ozone storage with Cloudera Data Warehouse Private Cloud](#).

September 16, 2022

There are no new features in the 1.4.0-H1 release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services. For a list of fixes, see [Fixed issues in Cloudera Data Warehouse Private Cloud \(1.4.0-H1\)](#).

June 21, 2022

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services introduces the new features and improvements that are described in this section.

Query isolation for scan-heavy, data-intensive queries in Hive LLAP Virtual Warehouses

Hive Virtual Warehouses base auto-scaling on the total scan size of the query. HiveServer, which receives all incoming queries, has a query planner component. When the HiveServer query planner receives queries, it examines the total scan size of each query. That is, it looks at the number of bytes read from the file system required to execute the query. If the Query Isolation feature has been enabled for a Virtual Warehouse and a query scans more data than the threshold set in the `hive.query.isolation.scan.size.threshold` parameter, the planner runs the query in isolation. This means that an isolated standalone executor group is spawned to run the data-intensive query. For more information, see [Hive query isolation for ETL jobs](#).

Generate and download Hive diagnostic bundles

You can now generate and download diagnostic bundles containing log files for Hive Virtual Warehouses. For more information, see [Downloading Hive diagnostic bundles in Data Warehouse Private Cloud](#).

SSL support added for MariaDB and MySQL databases

CDW can connect to SSL-enabled MariaDB and MySQL databases on the base cluster in addition to the SSL-enabled PostgreSQL database. For optimum security, the network connection between the default Database Catalog Hive

MetaStore (HMS) in CDW and the relational database hosting the base cluster's HMS must be encrypted with SSL. See [How to enable SSL support for MariaDB, MySQL, and Oracle databases](#).

SSL support added for Oracle databases (Preview)

CDW can connect to SSL-enabled Oracle database on the base cluster. See [How to enable SSL support for MariaDB, MySQL, and Oracle databases](#).



Note: Cloudera recommends that you explore this feature in development or test environments. SSL-enabled Oracle database is not recommended for use in production deployments.

Ability to configure Impala coordinator and executor pod size (Preview)

You can optimize the performance of your Impala Virtual Warehouse and resources used in an environment based on your hardware configuration by customizing the amount of resources allocated to the Impala coordinators, executors, and catalog daemons. This helps you to better leverage intra-query parallelism and achieve powerful compute clusters with fewer nodes. For more information, see [Creating custom pod configurations for Impala Virtual Warehouses](#).



Note: This is a preview feature and not recommended for use in production environments. Cloudera recommends that you explore this feature in test or development environments.

Node-level monitoring capabilities on ECS (Preview)

You can now monitor node-level metrics using Grafana. The metrics include CPU usage, memory usage, network usage, and disk IO. For more information, see [Monitoring Data Warehouse service resources with Grafana dashboards](#).



Note: This is a preview feature and not recommended for use in production environments. Cloudera recommends that you explore this feature in test or development environments.

Auto-shutdown Impala coordinators

When you create a Virtual Warehouse, you can configure Impala coordinators to automatically shutdown during idle periods. You can also set a delay before the coordinator shuts down. For more information, see [Configuring Impala coordinator shutdown](#).

Data Visualization integration in Cloudera Data Warehouse (Preview)

CDW integrates Data Visualization for building graphic representations of data, dashboards, and visual applications based on CDW data, or other data sources you connect to. Authorized users can explore data using graphics such as, pie charts and histograms and collaborate using dashboards. BI analysts who can access your environment can use these features. See [Creating a Data Visualization instance in CDW](#).



Note: This is a preview feature and not recommended for use in production environments. Cloudera recommends that you explore this feature in test or development environments.

Impala Debug Web UIs are available in CDW Private Cloud

In CDW Private Cloud, you can now use the Impala debug web UIs, which map to equivalent debug web UIs in Cloudera Manager as follows:

CDW Debug Web UI	Cloudera Manager Equivalent
Impala Catalog Web UI	Catalog Server Web UI
Impala Coordinator Web UI	Impala Daemon Web UI
Impala StateStore Web UI	StateStore Web UI

For more information about this feature, see [Using the Web UI to debug Impala Virtual Warehouses](#).

Hue supports Hive Hybrid Procedural SQL in CDW

You can run Hive Hybrid Procedural SQL (HPL/SQL) using the Hue query editor in CDW. For more information, see [How to run a stored procedure from Hue in Cloudera Data Warehouse](#).

Improved read performance of ORC tables by Impala

Continuous improvements in Impala's read performance of ORC tables.

Security improvements

- CVE-2021-44228 (Apache Log4j 2 vulnerability) has been addressed in CDW on CDP Private Cloud 1.4.0 by upgrading Apache Log4j 2 to version 2.17.1.
- CDW Docker containers, applications, and modules now run as non-root users, including Hue.

April 13, 2022

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services introduces the new features and improvements that are described in this section.

Support for SSL-enabled MariaDB and MySQL database (Preview)

CDW can connect to SSL-enabled MariaDB and MySQL databases on the base cluster. This feature is not ready for production deployment. Cloudera encourages you to explore this feature in a non-production environments and provide feedback on your experiences.



Note: You need to contact Cloudera to have this feature enabled.

Security improvements

CDW Docker containers, applications, and modules now run as non-root users. However, Hue containers still use root user.

January 14, 2022

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services introduces the new features and improvements that are described in this section.

CDW supports MySQL database for HMS on the base cluster

The Hive Metastore (HMS) on the CDP private base cluster supports MySQL databases in addition to PostgreSQL and MariaDB. For more information on database requirements, see [Base cluster database requirements for Cloudera Data Warehouse Private Cloud](#).

Users can access Hue using SSO

After logging in to the CDP Management Console, you can automatically log in to Hue using Single-Sign On (SSO). You do not need to enter your credentials to access Hue. To use this feature, you must first upgrade the CDW service to 1.3.3 (if you are on a previous version) and also create a new Virtual Warehouse. Upgrading an existing Virtual Warehouse to a newer version does not enable SSO authentication for Hue. The first user logging in to Hue will be a superuser.

CDW supports encrypted HDFS

CDW components such as Hive, Impala, and Hue can now read and write encrypted data stored on HDFS on the base cluster. If you are using Impala on an encrypted source data, then ensure that the data written to disk temporarily during processing is encrypted to keep the data confidential even during processing. For more information, see [Configuring Impala Virtual Warehouses to spill encrypted data to HDFS in CDW Private Cloud](#).

Support for using Hive user-defined functions (UDF) (Technical Preview)

You can export user-defined functions (UDF) to a JAR file from a Hadoop- and Hive-compatible Java project and store the JAR file on HDFS. Using Hive commands, you can register the UDF based on the JAR file, and call the UDF from a Hive query using Hue in CDW. For more information, see [Creating a user-defined function in Cloudera Data Warehouse](#).

Support for creating Impala tables in Kudu (Technical Preview)

You can configure an Impala Virtual Warehouse to connect to Kudu and create Impala tables in Kudu using the Hue query editor. For more information, see [Configuring Impala Virtual Warehouses to create Impala tables in Kudu in Cloudera Data Warehouse Private Cloud](#).

Support for using custom generated Kerberos keytabs (Technical Preview)

By default, the principals and keytabs in CDW are generated programmatically. You can generate and provide the keytabs, but the hostnames in the Kerberos principals are fixed. To learn more about this, see [Predefined Kerberos principals in Cloudera Data Warehouse Private Cloud](#).

CDW collects and provides diagnostic bundles for Impala

You can trigger a job to generate workload logs from the CDW web interface for Impala Virtual Warehouses. The diagnostic bundles are collated and stored under the tmp directory on HDFS. You can access and download the logs using the Hue service from the base cluster. For more information, see [Generating and downloading diagnostic bundles](#).

Security improvements

- CVE-2021-44228 (Apache Log4j 2 vulnerability) has been addressed in CDW on CDP Private Cloud 1.3.3 by upgrading Apache Log4j 2 to version 2.16.
- CDP now sets stricter file permissions for CDW secrets.

October 29, 2021

There are no new features in this release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services

October 4, 2021

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services introduces the new features and improvements that are described in this topic.

Embedded Container Service (ECS) support for Cloudera Data Warehouse (CDW)

ECS service creates and manages an embedded Kubernetes deployment for the CDW Data Service in your data center. Cloudera Manager handles the installation and configuration of ECS clusters while you manage and monitor your ECS clusters. For more information on ECS, see [ECS documentation](#).

Custom Security Context Constraint name in OpenShift environments

CDW provides an option to specify a custom Security Context Constraint (SCC) name while activating an OpenShift environment. This allows CDW to use 1000 as the UID for the Hive user. The custom SCC is not deleted when you deactivate or delete the CDW environment.

CDW supports Search and Bind users from multiple LDAP organizational units (OU)

Workload services such as Hue, Impala, and Hive now support LDAP Search Bind authentication with the following limitations, mostly related to filtering:

- You must specify at least two filter conditions, such as `(&(uid={0})(objectClass=person))`.
- You must specify at least two group search filter, such as `(&(memberUid={0})(objectClass=posixGroup)(gidNumber=4001))`.
- Hive CLI does not use the specified LDAP filter.
- Hue and Hive can use the User Search base for Group searches as well.
- Hue expects the groups to have 'cn' attribute as the group name attribute.
- Do not specify OU in the LDAP group search or user search. Remove “ou=xyz” from “ou=xyz,dc=cloudera,dc=com”, so that it reads “dc=cloudera,dc=com”.
- The group filter must be strict and must only return one group.
- Data Analytics Studio (DAS) can be configured only with Simple Bind authentication and uses the LDAP Bind DN as the Bind pattern.

Changes to supported databases for Hive Metastore on base cluster (Technical Preview)

The Hive Metastore (HMS) on the CDP Private Base cluster supports MariaDB databases in addition to PostgreSQL. This feature is in Technical Preview and is not ready for production deployment. For more information on database requirements, see [Base cluster database requirements for Cloudera Data Warehouse Private Cloud](#).

Ability to configure database name for Hue and DAS

You can specify a custom database name for Hue, and DAS during environment activation when creating a default Database Catalog with external databases, or while creating a non-default Database Catalog. For more information, see [Adding a new Database Catalog](#).



Note: If you are using an embedded database, then these fields do not apply.

DAS is disabled when HMS on the base cluster is configured with a non-PostgreSQL database

DAS supports only PostgreSQL databases. If you have configured your base cluster with a non-PostgreSQL database and if your Virtual Warehouse is linked to a default Database Catalog, then DAS is not installed on the cluster.

Ability to specify Delegation Username and Password

CDW no longer inherits the delegation user from the LDAP Bind DN. You must specify Delegation Username and Delegation Password on the Activation Settings page while creating a CDW environment. The delegation user and password can authenticate users through an LDAP service account. For more information, see [Using delegation username and password](#).

April 28, 2021

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud introduces the new features and improvements that are described in this topic.

New features

- Support for Red Hat OpenShift version 4.6
- Low resource hardware requirements mode for OpenShift environments: In this release, you can deploy CDW Private Cloud using fewer nodes, memory, and processors to get the functional benefits of CDW on Private Cloud. For details, see [Get started with OpenShift low resource mode requirements](#).



Note: Cloudera recommends that you use the Low Resource Mode option for proof of concept (POC) purposes only. This feature is not recommended for production deployment.

- Impala Virtual Warehouses now support ACID reads on ORC tables: Impala Virtual Warehouse users can now run SQL queries on ORC ACID tables.
- Hive ACID compaction fixes: Multiple stability fixes have been made to increase the efficiency of Hive ACID compaction operations.

Behavioural changes

- New hardware requirements for standard resource mode: For details, see [Get started with OpenShift standard resource mode requirements](#)

- Custom service accounts can be used when installing CDP Private Cloud 1.2 with the following limitations and workarounds:
 - Limitation (DWX-6845): Demo data fails to load.
 - Limitation (DWX-6954): Only Hue is supported as the SQL editor for both Hive and Impala Virtual Warehouses.
 - Workaround (DWX-5941, DWX-7152): When you use custom service accounts during CDP Private Cloud 1.2 installation, you must add the custom service account users to the HDFS, HBase, and Hadoop SQL services using the Ranger Admin Web UI in Cloudera Manager:

1. In Cloudera Manager, select **RangerRanger Admin Web UI** to launch the Service Manager page.
2. In the Service Manager page, select the edit icon for HDFS to edit the service:
 - a. In the Add New Configurations region of the Edit Service page, add the custom HDFS service account user to the Value listed for the following parameters:



tag.download.auth.users

policy.download.auth.users

Make sure you separate user names with a comma.

For example:

Add New Configurations

Name	Value	
tag.download.auth.users	hdfs, hdfs-custom-user	
policy.download.auth.users	hdfs, hdfs-custom-user	

Add the custom service user to both configuration parameters.

- b. Click Save to save your changes and return to the Service Manager page.
3. On the Service Manager page, click the edit icon for HBASE to edit the service:
 - a. In the Add New Configurations region of the Edit Service page, add the custom HBase service account user to the following configuration parameters:

policy.grantrevoke.auth.users

tag.download.auth.users

policy.download.auth.users

Separate user names with a comma.

For example:

Add New Configurations

Add the custom HBASE user to these configuration parameters:

Name	Value	
setup.additional.default.policies	true	✕
default-policy.1.resource.table	atlas_janus,ATLAS_ENTITY_AUDIT	✕
default-policy.1.resource.column-fa	*	✕
policy.grantrevoke.auth.users	hbase, hbase-custom-user	✕
default-policy.1.policyItem.1.acces	create,read,write,admin	✕
tag.download.auth.users	hbase, hbase-custom-user	✕
default-policy.1.name	ATLAS_HBASE_TABLES	✕
policy.download.auth.users	hbase, hbase-custom-user	✕
default-policy.1.policyItem.1.users	atlas	✕
default-policy.1.resource.column	*	✕

- b. Click Save to save your changes and return to the Service Manager page.
4. In the Service Manager page, under HBASE, click the cm_hbase link to navigate to the configuration page for cm_hbase policies:



5. On the cm_hbase Policies page, click the edit icon for the all-table, column-family, column policy and add the custom HBase service account user to this policy:

List of Policies : cm_hbase

Policy ID	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
4	all - table, column-family, column	--	Enabled	Enabled	--	--	hbase rangerlookup	✕
5	ATLAS_HBASE_TABLES	--	Enabled	Enabled	--	--	atlas	✕

6. Click Save to save your changes and navigate back to the Service Manager page.
7. In the Service Manager page, click the edit icon for HADOOP SQL to edit the service:
 - a. In the Edit Service page, scroll down to the Add New Configurations region and add the custom Impala and Hive service account users, separating user names with commas, to the following configuration parameters:
 - tag.download.auth.users
 - policy.download.auth.users
 - policy.grantrevoke.auth.users

- b. To the policy.download.auth.users also add {USER}, separating user names with commas, to configure access to your SQL editor.
- c. Click Save to save your changes.

December 16, 2020

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud introduces the new features and improvements that are described in this topic.

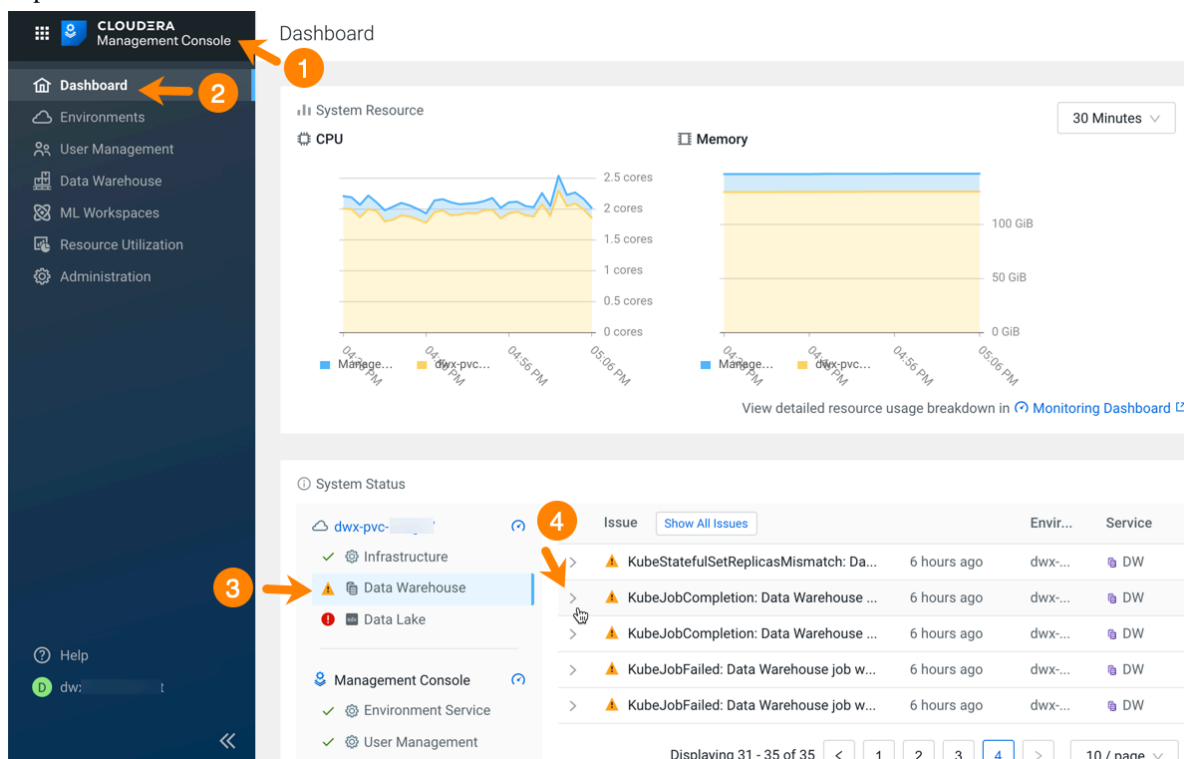
This release of CDW Private Cloud contains the following features:

- Support for Red Hat OpenShift version 4.5.
- Support for OpenLDAP in addition to Microsoft Active Directory LDAP.
- CDW endpoints on OpenShift clusters are now secured with TLS.

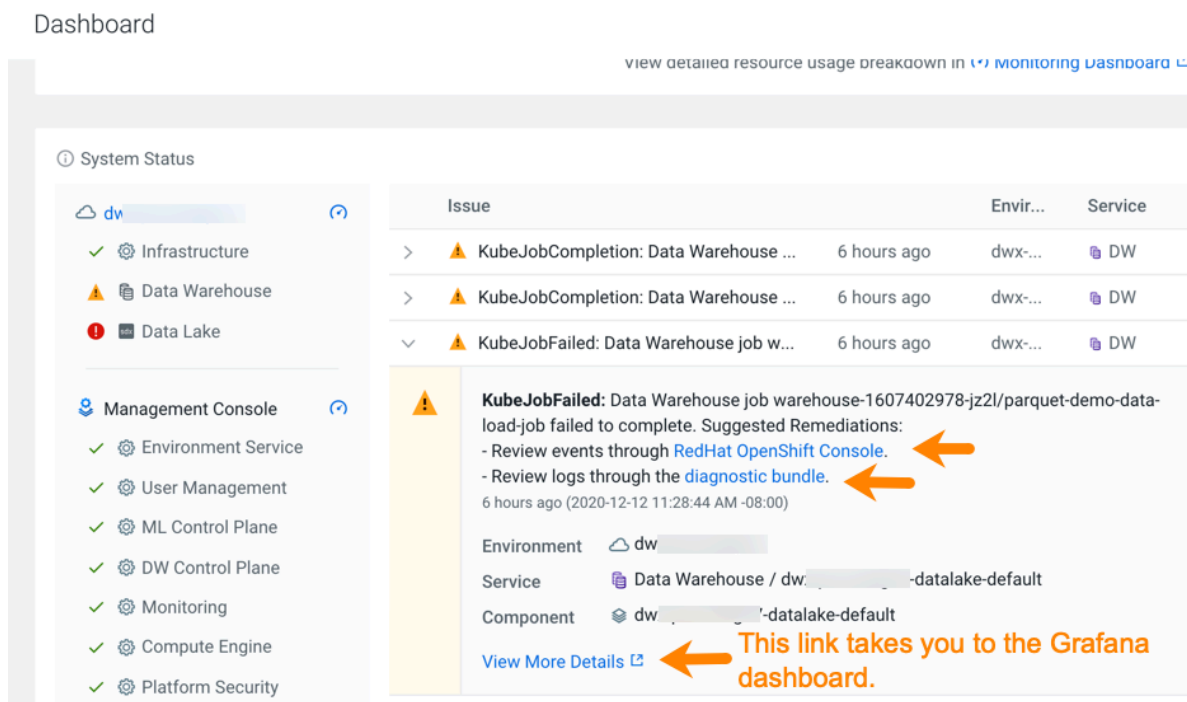


Tip: If you copied a URL to configure a connection in CDW Private Cloud 1.0, you must re-copy the new TLS-enabled URL in CDW Private Cloud 1.1. For example, you might have copied a URL to configure a connection in Data Analytics Studio, Hue, the JDBC URL, or a URL to connect with Impala shell. For information about managing Virtual Warehouse endpoints see, [SSL-enabled client endpoints](#).

- Alerts have been added for the CDW service in the Management Console Dashboard to monitor the service health in terms of resource usage, network/security issues, and pod issues:
1. Navigate to Management Console Dashboard Data Warehouse , and click the arrow adjacent to the alert to expand it:



2. After expanding the alert, there are multiple options for troubleshooting. For example, in this KubeJobFailed alert for Data Warehouse, you can review the event in RedHat OpenShift Console, select the diagnostic bundle link to download logs, or click View More Details to launch the Grafana dashboard for this issue:



August 17, 2020

This release of Cloudera Data Warehouse (CDW) service on CDP Private Cloud introduces the new features and improvements that are described in this topic.

This is the first release of CDW service Private Cloud. It contains the following features:

- "Noisy neighbor" prevention achieved through workload isolation.
- Hive and Impala Virtual Warehouses with concurrency, auto-scaling, and auto-suspend features.
- Data Analytics Studio (DAS) and Hue SQL to support querying and analysis.
- On-premises LDAP integration and SSL with PostgreSQL.
- Grafana monitoring dashboards for Hive and Impala.
- Diagnostics and log collection.
- Separate compute and storage on-premises.
- Improved CPU and memory utilization across the data warehousing apps.
- A central console from which you can secure and govern your data across data warehouses.

For an overview of CDW service Private Cloud and how to get started with CDW Private Cloud see [Private cloud support](#) and [Getting Started with OpenShift requirements](#).

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 in 1.4.1


DWX-13813: Unable to select or delete custom pod configurations for Impala

You may not be able to select an Impala pod configuration from the EDIT POD CONFIGURATIONS tab on the **Environment Details** page intermittently.

Also, you may not be able to delete a pod configuration if you click Apply or Apply Changes after it has been created and the "Configuration update initiated" message is displayed.

You can delete a pod configuration right after creating it, but before clicking Apply or Apply Changes.

DWX-13816: Refresh option at the environment level is not functional

The Refresh option on the more options menu () is clickable and shows an "Refresh of environment initiated" message too, but it does not refresh the data in the backend.

DWX-13759: DAS server and DAS WebApp pods exists even after disabling DAS

After you disable DAS from the **Advanced Settings** page, you may still see the DAS server and DAS WebApp pods running in the Database Catalog and Virtual Warehouse namespaces respectively.

Delete and recreate the Database Catalog and Virtual Warehouse.

Known Issues before 1.4.1

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

If the `/etc/krb5.conf` file on the Cloudera Manager host contains `"include"` or `"includedir"` directives, then you may encounter Kerberos-related failures on both Embedded Container Service and Red Hat OpenShift platforms. You may see the following error in the Database Catalog's metastore pod logs: `SERVICE_PRINCIPAL is required for kinit`.

To resolve this issue:

1. SSH into the Cloudera Manager host as an administrator.
2. Open the `/etc/krb5.conf` file for editing.
3. Comment the lines containing `"include"` and `"includedir"` directives.
4. Save the changes and exit.
5. Recreate the environment.

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-8525 and DWX-8526: CDW does not support upgrades from CDW 1.2 to 1.3.2

You may encounter issues while creating a new Virtual Warehouse on an existing Database Catalog (version 1.2) in an environment that is upgraded from 1.2 to 1.3.1.

The existing Virtual Warehouses will continue to operate. But if you want to create a new Virtual Warehouse, then you must reactivate the CDW environment after upgrading the base cluster from 1.2 to 1.3.2.

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.

Known issues on OpenShift cluster environments

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

Known issues in 1.4.1

No new known issues in this area.

Known Issues before 1.4.1

DWX-4723: HTTP 504 Gateway timeout error returned when you delete an environment

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

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

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

For more information about setting this timeout, see the [OpenShift documentation](#).

ECS cluster environments

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

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 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.

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 before 1.4.1

DWX-14014 (Affects 1.4.0): Hue Query Processor port is hardcoded

The Hue Query Processor port is hardcoded to 5432, which does not allow the connection to an external PostgreSQL database using custom ports. Database Catalogs fail to start when you configure custom port for the external database during installation.

Use the default port (5432) or use the embedded database for the Database Catalog.

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 in 1.4.1

DWX-13718: Unable to register UDFs in CDW using the auxiliary JARs approach

You may not be able to register UDFs by setting the `CDW_HIVE_AUX_JARS_PATH` parameter by going to Hive Virtual Warehouse Edit CONFIGURATIONS Hiveserver2 tab and selecting env from the Configuration file drop-down menu and repeating the same from the Query Coordinator tab.

You can register UDFs on Hiveserver by changing the volume mount of `init-aux-jar-localizer` init container from `hadoop-conf-final` to `hadoop-conf-volume` in `hiveserver2` statefulset using Kubernetes IDE or `kubectl`.

DWX-13095: Hive LLAP coordinator and executor not getting up

You may notice that Hive query executor pods disappear every 10 minutes after a spike of connections in ZooKeeper. This happens because the ZooKeeper node (/zkdtsm_hive-[/****COMPUTE-NAMESPACE****]/ZKDTSMRoot/ZKDTSMTokensRoot) that stores the delegation tokens increases in size because a number of delegation tokens are not cleaned up. As a result, it takes more than a minute for the ZooKeeper cache to rebuild, slowing down the query executor startup. This slowness in starting up the query executors results into a time out causing the query executors to disappear and start up again every 10 minutes.

You can use one of the following workarounds to resolve this issue:

- Create a new Virtual Warehouse, or
- Scale down the executors to zero, clean up the delegation tokens in ZooKeeper in the /zkdtsm_hive-[/****COMPUTE-NAMESPACE****]/ZKDTSMRoot/ZKDTSMTokensRoot znode path, and then scale back the executors, or
- Upgrade to a CDP Private Cloud Data Services release 1.5.0 or higher

Known Issues before 1.4.1

DWX-8524: Hive queries on sys.db tables may fail

When you run queries on the sys.db table from Data Analytics Studio, you may see the following error: “Cannot create PoolableConnectionFactory (Could not open SSL root certificate file /mnt/config/current/hms-root.crt.)” on a default Database Catalog using PostgreSQL database. This happens because PostgreSQL requires SSL. This issue does not occur on Database Catalogs using MariaDB.

None.

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

The slider that can be used to set WAIT TIME in the Hive Virtual Warehouse UI displays a range from 0 to 1000, but if you set it to 0, the UI automatically changes it to 60.

None.

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 in 1.4.1

No new known issues in this area.

Known Issues in 1.4.0

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

Hue and Impala shell display a 401 Unauthorized error when you submit a query to an Impala Virtual Warehouse. This issue can happen if the user who is trying to access Impala is not a part of an LDAP group.

Add the user to an LDAP group. Users must be part of at least one LDAP group for Impala.

DWX-10382: Unable to connect to the pre-upgrade Impala Virtual Warehouse using impala-shell on an Embedded Container Service (ECS) cluster

After upgrading from 1.3.2, 1.3.3, or 1.3.4 to 1.4.0, you may not be able to access the pre-upgrade Impala coordinators through its endpoints.

Restarting the coordinator pods can solve the issue. Alternatively, you can patch the leader elector image in Impala coordinators after upgrading CDW as follows. This is necessary because the old leader elector image runs deprecated code which is not compatible with the latest RKE versions:

1. Start a terminal session that can connect to your Kubernetes cluster using kubeconfig, for example.
2. Run the following script:

```
#!/usr/bin/env bash
set -x

DWX_VERSION="1.4.0-b26"
NAMESPACE=${NAMESPACE:?please set the NAMESPACE to update}
LEADER_ELECTOR_IMAGE_PRE_UPGRADE=$(kubectl get sts coordinator -n $NAMESPACE \
-o jsonpath={.spec.template.spec.containers[3].image})

LEADER_ELECTOR_IMAGE_POST_UPGRADE=$(sed \
"s/.*:.*:$DWX_VERSION/" <<< $LEADER_ELECTOR_IMAGE_PRE_UPGRADE)
kubectl patch sts coordinator -n $NAMESPACE --type='json' -p='[{"op": "replace", "path": "/spec/template/spec/containers/3/image", "value":"' $LEADER_ELECTOR_IMAGE_POST_UPGRADE' "}]'

kubectl delete pod -l app=coordinator
```

Known issues in Hue

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

Known Issues in 1.4.1

No new known issues in this area.

Known Issues 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.

Data Analytics Studio

This topic describes the Data Analytics Studio (DAS) known issues for Cloudera Data Warehouse Private Cloud.

Known Issues in 1.4.1

No new known issues in this area.

Known Issues before 1.4.1

DWX-9034: Query search page does not display the query history

The default value of the pod PID limit setting on Openshift is 1024. The das-event-processor pod in the Data Catalog's namespace may run out of the available processes over time on OpenShift, causing the query search page to not display query history. You may also see the following

exception in the Event Processor logs: "ERROR com.hortonworks.hivestudio.eventProcessor.meta.MetaInfoUpdater - Error occurred while reading dump metadata null".

You must manually restart the pod to resolve this issue. Alternatively, you can increase the pod PID limit or set it to maximum, which is a node level setting in Kubernetes/Openshift.

Fixed issues in Cloudera Data Warehouse Private Cloud

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

DWX-12846: Impala uses {0} and {1} incorrectly in group search filters

User group resolution was impacted because of the way in which Impala was configured to use the {0} and {1} placeholder strings in the LDAP configuration. This caused authorization issues for the users who were trying to connect to the Virtual Warehouse. This issue has been fixed.



Note: If you had changed the placeholder strings as a workaround, then you must change the values back to {0} and {1} for username and distinguished name (DN) respectively.

DWX-12613: Diagnostic bundle collection for Hive does not support custom time intervals and log types

Setting the time interval and the log types for generating the Hive diagnostic bundles works as expected and the output now can contain log files based your selection.

DWX-12468: Query store not reachable

With the addition of the Query Processor service, you can now see the Queries tab on the **Job Browser** page, along with the query history.

DWX-9245: Demo data set does not load on containerized clusters

Earlier, on containerized clusters, the job to load the demo data use to fail when you create a Database Catalog with the Load Demo Data option and the demo data set was not available in Hue. This issue has been fixed.

DWX-5496: If Cloudera Manager truststore or the external database/vault is updated, CDW environment activation fails

This issue is no longer present because the new refresh workflow systematically updates the certificates and database credentials.

DWX-12801: Authorization error on SSO login in Cloudera Data Visualization on CDW Private Cloud

Cloudera Data Visualization displayed "Access Denied" error when you logged in using SSO. This issue has been resolved.

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 1.4.1 release:

Runtime component	Version
Hive	2022.0.10.1-8
Impala	2022.0.10.1-8
Hue	2022.0.10.1-8