

Release Notes

Date published:

Date modified:

CLOUDERA

Legal Notice

© Cloudera Inc. 2025. All rights reserved.

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

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

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

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

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

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

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

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

Contents

Overview.....	5
Cloudera Runtime Component Versions.....	5
What's New In Cloudera Runtime 7.1.0.0.....	6
What's New in Apache Atlas.....	6
What's new in DAS.....	6
What's New in Apache HBase.....	6
What's New in Apache Hadoop HDFS.....	6
What's New in Apache Hive.....	6
What's New in Hue.....	7
What's New in Apache Impala.....	8
What's New in Apache Kafka.....	9
What's New in Apache Knox.....	9
What's New in Apache Kudu.....	10
What's New in Apache Oozie.....	10
What's New in Apache Phoenix.....	10
What's New in Cloudera Search.....	10
What's New in Apache Spark.....	10
What's New in Apache Hadoop YARN.....	10
What's New in Apache ZooKeeper.....	10
Fixed Issues In Cloudera Runtime 7.1.0.0.....	11
Atlas.....	11
Fixed issues in DAS.....	11
Hadoop.....	11
HDFS.....	12
Fixed Issues in Apache Hive.....	12
Solr.....	12
Zeppelin.....	12
Known Issues In Cloudera Runtime 7.1.0.0.....	12
Known Issues in Apache Atlas.....	13
Known Issues in DAS.....	15
Known Issues in Apache HBase.....	16
Known Issues in HDFS.....	28
Known Issues in Apache Hive.....	29
Known Issues in Hue.....	31
Known Issues in Apache Impala.....	35
Known Issues in Apache Kafka.....	40
Known Issues in Kerberos.....	42
Known Issues in Apache Knox.....	42
Known Issues in Apache Kudu.....	43
Known Issues in Apache Oozie.....	43
Known Issues in Apache Ranger.....	43

Known Issues in Cloudera Search.....	44
Known Issues in Apache Solr.....	47
Known Issues in Apache Spark.....	47
Known Issues for Apache Sqoop.....	48
Known Issues in MapReduce and YARN.....	48
Known Issues in Apache Zeppelin.....	50
Known Issues in Apache ZooKeeper.....	51

Overview

Welcome to the Cloudera Runtime Release Notes. This document provides you with the latest information about Cloudera Runtime 7.1.0.0. This document includes improvements and describes new features, bug fixes, tech previews and more. For detailed information about the runtime components themselves, see [Cloudera documentation](#).

Cloudera Runtime Component Versions

List of the official Apache component versions for Cloudera Runtime. To know the Apache component versions for compatibility with other applications, you must be familiar with the latest Apache component versions in Cloudera Runtime. You should also be aware of the available Technical Preview components and use them only in a testing environment. Apache versions of Cloudera Runtime 7.1.0.0. components.

Component	Apache Version
Apache Atlas	2.0.0
Apache Arrow	0.8.0
Apache Avatica	1.10.0
Apache Avro	1.8.2
Apache Calcite	1.19.0
Apache Crunch	0.11.0
Apache DataFu	1.3.0
Apache Druid	0.15.1
Apache Hadoop	3.1.1
Apache HBase	2.2.0
HBase Indexer	1.5.0
Apache Hive	3.1.3000
Hue	4.5.0
Apache Impala	3.3.0
Apache Kafka	2.3.0
Apache Knox	1.3.0
Apache Kudu	1.11.0
Apache Livy	0.6.0
Apache Oozie	5.1.0
Apache ORC	1.5.1
Apache Parquet	1.10.99
Apache Phoenix	5.0.0
Apache Ranger	2.0.0
Apache Solr	8.4.1
Apache Spark	2.4.0
Apache Sqoop	1.4.7
Apache Superset	0.34.0
Apache Tez	0.9.1

Component	Apache Version
Apache ZooKeeper	3.5.5
Apache Zeppelin	0.8.2
YARN	3.1.1

What's New In Cloudera Runtime 7.1.0.0

This version of Cloudera Runtime provides you with several new capabilities. Learn how the new features and improvements benefit you.

What's New in Apache Atlas

This topic lists new features for Apache Atlas in this release of Cloudera Runtime.

Associated Classification Counts and Search Links in Statistics

Atlas Statistics page now includes a list of the classifications that have been associated with entities. For each classification, Atlas shows a count of the associated entities; clicking the count opens search results for a search on that classification.

See [Apache Atlas Statistics reference](#)

New Atlas Metrics for Cloudera Manager Charts

Atlas now provides a number of metrics that you can use to create time-series charts in Cloudera Manager. Add these charts to a dashboard in Cloudera Manager to see Atlas resource use at a glance.

Metrics include the rate of entity creation, update, and deletion and the size of Java heap used by Atlas.

See [Charting Time-Series Data](#)

What's new in DAS

There are no new features in DAS in this release of Cloudera Runtime.

What's New in Apache HBase

There are no new features for Apache HBase in this release of Cloudera Runtime.

For more information about Apache HBase, see [Apache HBase Overview](#).

What's New in Apache Hadoop HDFS

There are no new features for Apache Hadoop HDFS in this release of Cloudera Runtime.

For more information about HDFS, see [HDFS Overview](#)

What's New in Apache Hive

There are no new features for Apache Hive in this release of Cloudera Runtime.

What's New in Hue

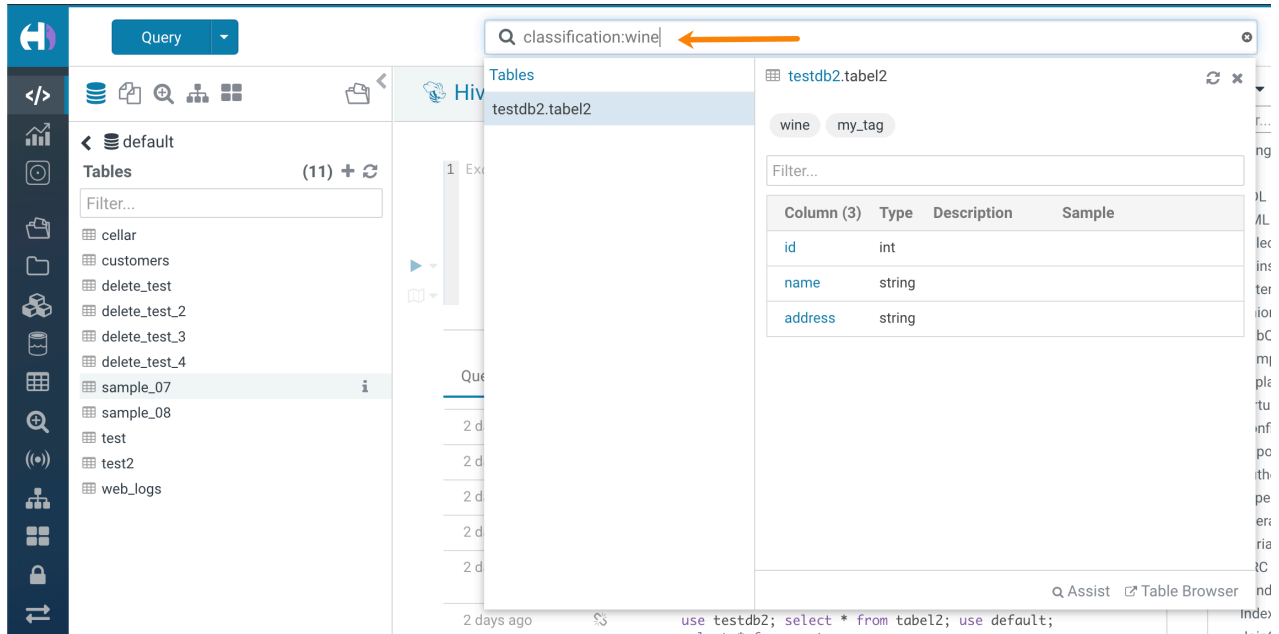
This topic lists new features for Hue in this release of Cloudera Runtime.

Support for Hive on Tez

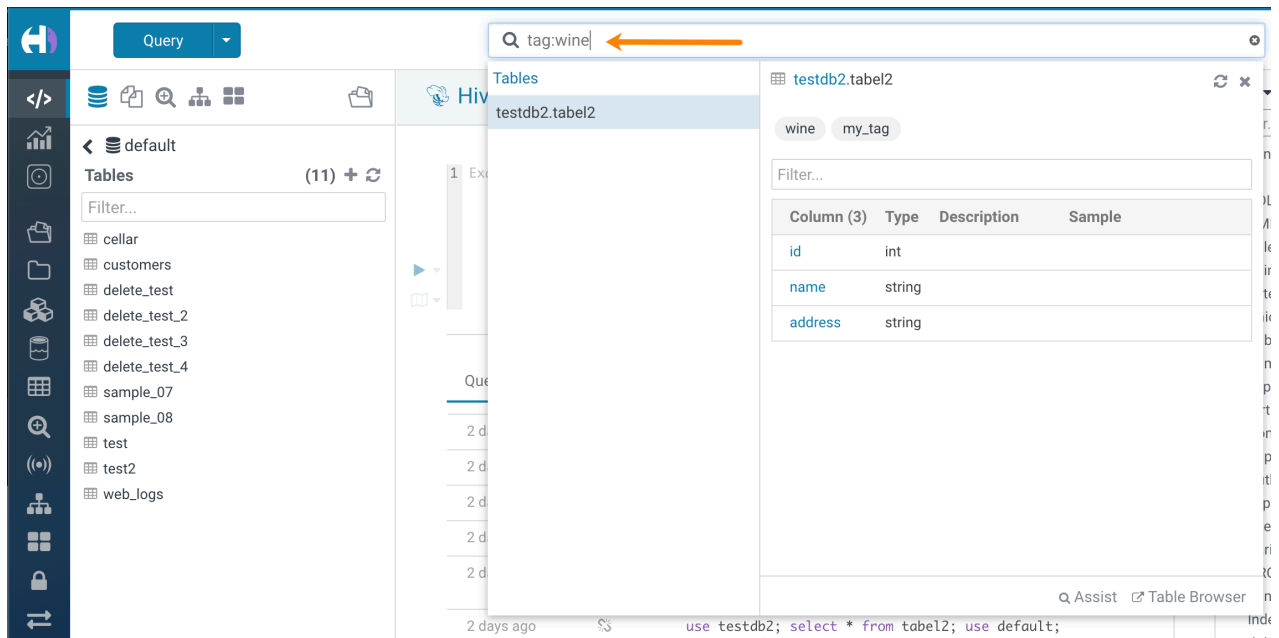
Hue in Cloudera Runtime now supports Hive using Tez as its execution engine.

Integration with Apache Atlas data catalog

You can now locate tables by searching for Atlas classifications in Hue by specifying the classification search term in the search box at the top of the page:

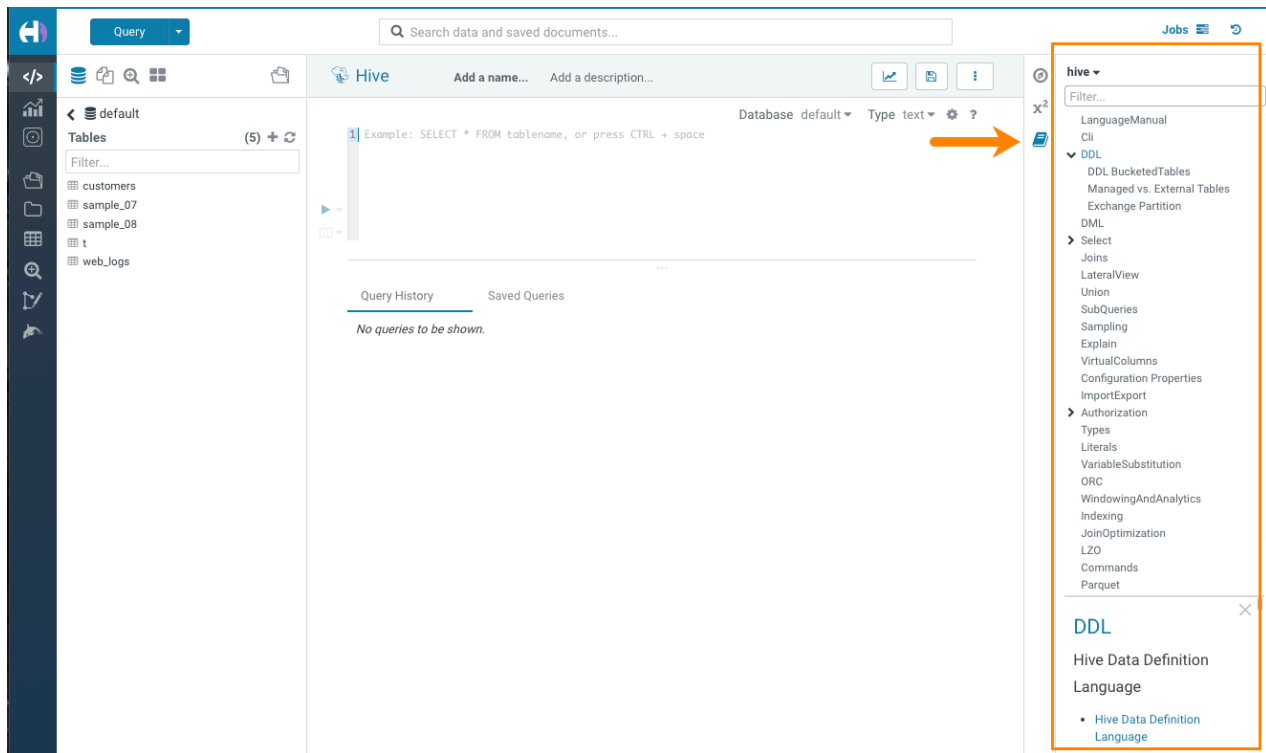


You can also locate tables by searching for Atlas classifications by specifying the tag search term:



Hive Language Reference

A language reference has been added for Hive:



When you select the Hive query editor, click the book icon  to the right of the editor window to launch the Hive language reference.

What's New in Apache Impala

This topic lists new features for Apache Impala in this release of Cloudera Runtime.

April, 2020

Increased scratch capacity

To help reduce spilling to disk:

- Added startup parameter to support Spill-to-disk compression to increase effective scratch capacity by 2.5x.
- Added startup parameter to reclaim space in scratch files.

Improved data cache performance

- Improved the efficiency of the data cache by providing an option to use a different cache eviction algorithm (LIRS).

Support for Kudu Date and Varchar column types

Support for reading ZSTD-compressed text files

For details, see [Using Text Data Files](#).

Improved read performance of ORC tables

Improved Impala resiliency

This release adds client retry support in the impala-shell. For details about installing the impala-shell, see [Connecting to Impala Virtual Warehouse from Impala shell](#).

March, 2020

broadcast_bytes_limit query option

In this release, you can set a limit for the size of a broadcast input. For details, see [Impala Query Options](#).

ORC stability and performance improvements

ORC reads enabled by default

Impala stability and performance have been improved. Consequently, ORC reads are now enabled in Impala by default. To disable, set --enable_orc_scanner to false when starting the cluster.

February, 2020

Constraints

This release adds support for primary and foreign key constraints, but in this release the constraints are advisory and intended for estimating cardinality during query planning in a future release. There is no attempt to enforce constraints. For details, see the “Constraints” section of [Create Table Statement](#).

Enhanced external Kudu table

By default HMS implicitly translates internal Kudu tables to external Kudu tables with the 'external.table.purge' property set to true. These tables behave similar to internal tables. You can explicitly create such external Kudu tables. For details, see the “External Kudu Tables” section of [Create Table Statement](#).

Ranger column masking

This release supports Ranger column masking, which hides sensitive columnar data in Impala query output. For example, you can define a policy that reveals only the first or last four characters of column data. Column masking is enabled by default. For details, see the "Ranger Column Masking" section in [Impala Authorization](#).

What's New in Apache Kafka

This topic lists new features for Apache Kafka in this release of Cloudera Runtime.

Zookeeper TLS/SSL support for Kafka

A new Kafka configuration parameter, Enable Secure Connection to ZooKeeper is added. This property enables or disables secure communication between Kafka and Zookeeper. The property is enabled by default, however, it will only take effect if the Enable TLS/SSL for ZooKeeper property is also enabled for the dependent ZooKeeper service. For more information, see the related [Zookeeper documentation](#) and [Kafka documentation](#).

Enable Delegation Tokens is set to true by default

The Enable Delegation Tokens property that enables and disables delegation token based authentication is set to true by default. Note however, that this does not by default enable the feature itself as delegation tokens require Kerberos authentication to function.

For more information, see [Enable or Disable Authentication with Delegation Tokens](#).

What's New in Apache Knox

There are no new features for Apache Knox in this release of Cloudera Runtime.

What's New in Apache Kudu

This topic lists new features for Apache Kudu in this release of Cloudera Runtime.

Support for EC-based ciphers

Kudu supports EC-based ciphers in the web UI. Previously, both kudu-tserver and kudu-master failed to start if supplied with server certificates based on ciphers that squeasel does not support.

What's New in Apache Oozie

There are no new features for Apache Oozie in this release of Cloudera Runtime.

For more information about Oozie, see [Overview of Oozie](#).

What's New in Apache Phoenix

This release of Cloudera Runtime introduces Apache Phoenix, a massively parallel, relational database engine supporting OLTP use cases using Apache HBase as its backing store.

For more information about Apache Phoenix, see [Phoenix Overview](#).



Note: Apache Phoenix in CDP is a technical preview and considered under development. Do not use this in your production systems. If you have feedback, contact Support by logging a case on the Cloudera Support Portal at <https://my.cloudera.com/support.html>. Technical preview features are not guaranteed troubleshooting and fixes.

What's New in Cloudera Search

There are no new features for Cloudera Search in this release of Cloudera Runtime.

For more information about Cloudera Search, see [Cloudera Search Overview](#).

What's New in Apache Spark

This topic lists new features for Apache Spark in this release of Cloudera Runtime.

Data engineering cluster

You can create a data engineering cluster in Amazon AWS from within CDP by selecting the Data Engineering cluster template. A data engineering includes Spark, Livy, Hive, Zeppelin, and Oozie, along with supporting services (HDFS, YARN, and Zookeeper).

See [Creating a Cluster on AWS](#).

What's New in Apache Hadoop YARN

There are no new features for Apache Hadoop YARN in this release of Cloudera Runtime.

For more information about Apache Hadoop YARN, see [Apache Hadoop YARN Overview](#)

What's New in Apache ZooKeeper

This topic lists new features for Apache ZooKeeper in this release of Cloudera Runtime.

ZooKeeper TLS/SSL

TLS/SSL encryption between the ZooKeeper client and the ZooKeeper server and within the ZooKeeper Quorum is supported. TLS/SSL encryption is automatically enabled when AutoTLS is enabled, or you can enable it manually using Cloudera Manager. ClientSSL requires you to enable and configure it by other components that you want to use this feature with. For more information, see [Configure ZooKeeper TLS/SSL using Cloudera Manager](#).

Fixed Issues In Cloudera Runtime 7.1.0.0

Fixed issues represent issues reported by Cloudera customers that are addressed in this release.

Atlas

This section lists the issues that have been fixed since the previous version.

OPSAPS-53788: Lineage info for HBase not available in a TLS-enabled environment

Atlas will not collect metadata from HBase (namespaces, tables, or column families) when Cloudera Manager is configured for TLS/SSL. This issue has no effect on HBase functionality, however, tag-based authorization through Ranger policies will not be available for HBase.

CDPD-6023: Slight differences in metadata collected for Hive database entities depending on source

When a Hive database is created using Beeline, the ownerType is sent to ATLAS_HOOK and is recorded in the hive_db entity. However, when a Hive database is created through Spark-shell, the ownerType is missing in the Atlas entity.

OPSAPS-51224: Atlas custom properties ignored in client services

When adding a custom property for the atlas-application.properties in Atlas hook-based services such as Hive, HBase, and Impala, the custom property is not reflected in the actual configuration file that Cloudera Manager generates, causing these properties to be ignored.

CDPD-5314: Unexpected Results from Atlas Advanced Search Query with Underscore Character

An Atlas Advanced Search with an equal operator (=) and a like operator where the like object includes an underscore does not return correct results. For example, the following search does not return expected results:

```
db.name : "=" finance and QualifiedName : "like" *finance_09*
```

CDPD-5309: Unexpected Results from Atlas Advanced Search Query on String Attributes

When new entity types are defined with string type attributes, Advanced Searches against the string attributes return unexpected results. For example, if attribute values include "caterpillar", "alleycat", and "altercation", all values are returned when searching for "cat*". The expected result would be that only "caterpillar" would be returned.

Fixed issues in DAS

This section lists the issues that have been fixed since the previous version.

- Fixed several stability issues in the reporting pipeline which caused the report pipeline to get stuck.
- Fixed performance issue in the events pipeline when events are on the S3 filesystem.
- [DWX-2508](#): DAS now shows accurate compilation time for a query on the Query Details Timeline tab.
- [DWX-2220](#): The **Compose** page UI has been enriched with tooltips and also displays the database count.

Hadoop

This section lists the issues that have been fixed since the previous version.

- [CDPD-7383](#): The ABFS and ADL connectors compatible with Alpine Linux and other platforms which have libssl1.1-1.1.1b-r1 as their native OpenSSL implementation, see [HADOOP-16460](#) and [HADOOP-16438](#)

HDFS

This section lists the issues that have been fixed since the previous version.

CDPD-6100

This improvement makes HDFS NameNode leave safemode immediately if all blocks have reported in, reducing the cluster startup time.

CDPD-2946: Slow reading and writing of erasure-coded files

The ISA-L library is not packaged with HDFS as a result of which HDFS erasure coding falls back to the Java implementation which is much slower than the native Hadoop implementation. This slows down the reading and writing of erasure-coded files.

Fixed Issues in Apache Hive

This section lists the issues that have been fixed since the previous version.

- [CDPD-676](#): Generate Oozie workflow for microstrategy
- [CDPD-7963](#): Fixes memory leak in HiveServer2 caused by external or unmanaged sessions in Tez..

Solr

This section lists the issues that have been fixed since the previous version.

Technical Service Bulletins

TSB-2021 389: CVE 2019-17558: Remote Code Execution in Solr through Velocity templates

For the latest update on this issue see the corresponding Knowledge article: [TSB-2021 389: CVE 2019-17558: Remote Code Execution in Solr through Velocity templates](#)

Zeppelin

This section lists the issues that have been fixed since the previous version.

CDPD-1683: Zeppelin demo users have been removed

Use cluster users to access Zeppelin. For information on provisioning users in CDP, see [Onboarding users](#).

CDPD-880, CDPD-1685: Shell, JDBC, and Spark interpreters have been removed

Workaround: Use an available interpreter. For Spark functionality, use the Livy interpreter.

CDPD-3047: Markdown interpreter does not handle certain numbered list syntax correctly

Using the plus sign (+) or asterisk (*) to continue a numbered list using the %md interpreter results in bullet point entries instead.

Workaround: None.

Known Issues In Cloudera Runtime 7.1.0.0

Known Issues in Apache Atlas

This topic describes known issues and workarounds for using Atlas in this release of Cloudera Runtime.

Bridge importing metadata from HBase fails when it encounters an HBase table for which a column family was previously dropped. The error indicates:

Metadata service API org.apache.atlas.AtlasClientV2\$API_V2@58112bc4 failed with status 404 (Not Found) Response Body ({ "errorCode": "ATLAS-404-00-007", "errorMessage": "Invalid instance creation/updation parameters passed : hbase_column_family.table: mandatory attribute value missing in type hbase_column_family" })

Workaround: None.

Cloudera JIRA: CDPD-7982

Hive Default Database Location Incorrect in Atlas Metadata

The location of the default Hive database as reported through the HMS-Atlas plugin does not match the actual location of the database. This problem does not affect non-default databases.

Workaround: None.

Cloudera JIRA: CDPD-6042

Unexpected Search Results When Using Regular Expressions in Basic Searches on Classifications

When you include a regular expression or wildcard in the search criteria for a classification in the Basic Search, the results may differ unexpectedly from when full classification names are included. For example, the Exclude sub-classifications option is respected when using a full classification name as the search criteria; when using part of the classification name and the wildcard (*) with Exclude sub-classifications turned off, entities marked with sub-classifications are not included in the results. Other instances of unexpected results include case-sensitivity.

Workaround: None.

Cloudera JIRA: CDPD-5933, CDPD-5931

Spark metadata order may affect lineage

Atlas may record unexpected lineage relationships when metadata collection from the Spark Atlas Connector occurs out of sequence from metadata collection from HMS. For example, if an ALTER TABLE operation in Spark changing a table name and is reported to Atlas before HMS has processed the change, Atlas may not show the correct lineage relationships to the altered table.

Workaround: None.

Cloudera JIRA: CDPD-4762

Searches for Qualified Names with "@" doesn't fetch the correct results

When searching Atlas qualifiedName values that include an "@" character (@), Atlas does not return the expected results or generate appropriate search suggestions.

Workaround: Consider leaving out the portion of the search string that includes the @ sign, using the wildcard character * instead.

Cloudera JIRA: CDPD-4545

Missing Impala and Spark lineage between tables and their data files

Atlas does not create lineage between Hive tables and their backing HDFS files for CTAS processes run in Impala or Spark.

Workaround: None.

Cloudera JIRA: CDP-5027, CDPD-3700, IMPALA-9070

Table alias values are not found in search

When table names are changed, Atlas keeps the old name of the table in a list of aliases. These values are not included in the search index in this release, so after a table name is changed, searching on the old table name will not return the entity for the table.

Workaround: None.

Cloudera JIRA: CDPD-3208

Hive lineage missing for INSERT OVERWRITE queries

Lineage is not generated for Hive INSERT OVERWRITE queries on partitioned tables. Lineage is generated as expected for CTAS queries from partitioned tables.

Workaround: None.

Cloudera JIRA: CDPD-3160

Logging out of Atlas does not manage the external authentication

At this time, Atlas does not communicate a log-out event with the external authentication management, Apache Knox. When you log out of Atlas, you can still open the instance of Atlas from the same web browser without re-authentication.

Workaround: To prevent access to Atlas after logging out, close all browser windows and exit the browser.

Cloudera JIRA: CDPD-3125

Ranking of top results in free-text search not intuitive

The Free-text search feature ranks results based on which attributes match the search criteria. The attribute ranking is evolving and therefore the choice of top results may not be intuitive in this release.

Workaround: If you don't find what you need in the top 5 results, use the full results or refine the search.

Cloudera JIRA: CDPD-1892

Free text search in Atlas is case sensitive

The free text search bar in the top of the screen allows you to search across entity types and through all text attributes for all entities. The search shows the top 5 results that match the search terms at any place in the text (*term* logic). It also shows suggestions that match the search terms that begin with the term (term* logic). However, in this release, the search results are case-sensitive.

Workaround: If you don't see the results you expect, repeat the search changing the case of the search terms.

Workaround: None.

Cloudera JIRA: CDPD-1884

Queries with ? wildcard return unexpected results

DSL queries in Advanced Search return incorrect results when the query text includes a question mark (?) wildcard character. This problem occurs in environments where trusted proxy for Knox is enabled, which is always the case for CDP.

Workaround: None.

Cloudera JIRA: CDPD-1823

Guest users are redirected incorrectly

Authenticated users logging in to Atlas are redirected to the CDP Knox-based login page. However, if a guest user (without Atlas privileges) attempts to log in to Atlas, the user is redirected instead to the Atlas login page.

Workaround: To avoid this problem, open the Atlas Dashboard in a private or incognito browser window.

Cloudera JIRA: CDPD-1664

All Spark Queries from the Same Spark Session are Included in a Single Atlas Process

A Spark session can include multiple queries. When Atlas reports the Spark metadata, it creates a single process entity to correspond to the Spark session. The result is that an Atlas lineage picture may show multiple input entities or multiple output entities from a process that are only related by the fact that they were included in operations in the same Spark session. The consequence of this behavior is that classifications will be propagated from any input entity to all output entities, even if the output entities aren't derived from the input entity.

Workaround: you can manually stop classification propagation to the inappropriate entity or choose not to propagate classifications that might be used in these scenarios.

Cloudera JIRA: CDPD-372

Known Issues in DAS

This topic describes known issues and workarounds for using DAS in this release of Cloudera Runtime.

- The task logs for a particular task may not be available in the task swimlane. And the zip file generated by download logs artifact may not have task logs, but instead contain an error-reports.json file with the error log of the download failures.
- You may not see any data for a report for any new queries that you run. This can happen especially for the last one day's report.

Workaround:

1. Shut down the DAS Event Processor.
2. Run the following command from the Postgres server:

```
update das.report_scheduler_run_audit set status = 'FAILED' where status = 'READING';
```

3. Start the DAS Event Processor.
- On clusters secured with Knox proxy only: You might not be able to save the changes to the JDBC URL in the DAS UI to change the server interface (HS2 or LLAP) on which you are running your queries.
 - You may be unable to upload tables or get an error while browsing files to upload tables in DAS on a cluster secured using Knox proxy.
 - DAS does not parse semicolons (;) and double hyphens (--) in strings and comments.

For example, if you have a semicolon in query such as the following, the query might fail: `select * from properties where prop_value = "name1;name2";`

If a semicolon is present in a comment, then execute the query after removing the semicolon from the comment, or removing the comment altogether. For example:

```
select * from test; -- select * from test;
select * from test; /* comment; comment */
```

Queries with double hyphens (--) might also fail. For example:

```
select * from test where option = '--name';
```

- You might face UI issues on Google Chrome while using faceted search. We recommend you to use the latest version of Google Chrome (version 71.x or higher).
- You are unable to see databases or the query editor on the **Compose** page even after following the troubleshooting procedure. This can be caused due to incorrect hive service name in ZooKeeper.
- Visual Explain for the same query shows different graphs on the **Compose** page and the **Query Details** page.
- While running some queries, if you restart HSI, the query execution is stopped. However, DAS does not reflect this change and the queries appear to be in the same state forever.

- After a fresh installation, when there is no data and you try to access the Reports tab, DAS displays an "HTTP 404 Not Found" error.
- Join count does not get updated for tables with partitioned columns.

Known Issues in Apache HBase

This topic describes known issues and workarounds for using HBase in this release of Cloudera Runtime.

HBASE-24885: If an operator uses HBCK2 to invoke multiple `assigns` operations against one Region or happens to invoke HBCK2 `assigns` while HBase is re-assigning a Region, it is possible that the Region will be abnormally assigned. For example, unassigned, stuck in transition, and doubly-assigned.

Obtain a fix for this issue. Operators should definitely not schedule multiple assigns for a single Region at the same time, however there is still a potential race condition.

OpDB Data Hub cluster fails to initialize if you are reusing a cloud storage location that was used by an older OpDB Data Hub cluster

Workaround: Stop HBase using Cloudera Manager before deleting an OpDB Data Hub cluster.

IntegrationTestReplication fails if replication does not finish before the verify phase begins

During IntegrationTestReplication, if the verify phase starts before the replication phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the -t flag to set the timeout value before starting verification.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

Workaround: N/A

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a Delete Table fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.
- If hbaseAdmin.modifyTable() is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The portOperation is implemented only for postDeleteColumn().
- If Create Table fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: N/A

Apache Issue: [HBASE-6992](#)

Bulk load is not supported when the source is the local HDFS

The bulk load feature (the completebulkload command) is not supported when the source is the local HDFS and the target is an object store, such as S3/ABFS.

Workaround: Use distcp to move the HFiles from HDFS to S3 and then run bulk load from S3 to S3.

Apache Issue: N/A

Storing Medium Objects (MOBs) in HBase is currently not supported

Storing MOBs in HBase relies on bulk loading files, and this is not currently supported when HBase is configured to use cloud storage (S3).

Workaround: N/A

Apache Issue: N/A

Known incompatibilities when upgrading from CDH and HDP

This section describes incompatible changes when you upgrade from CDH and HDP. Use this list to understand some of the deprecated items and incompatibilities if you are moving from HDP 2.x or CDH 5.x. Note that Cloudera Runtime uses HBase 2.x.x whereas older CDH and HDP used Apache HBase 1.x.



Important: Some APIs that are listed as deprecated, but these APIs do not block your upgrade. You must stop using the deprecated APIs in your existing applications after upgrade, and not use these APIs in new development.

List of Major Changes

- HBASE-16189 and HBASE-18945- You cannot open the HBase 2.x HFiles through HBase 1.x version.
- HBASE-18240 - Changed the ReplicationEndpoint Interface. It also introduces a new hbase-third party 1.0 that packages all the third party utilities, which are expected to run in the hbase cluster.

Co-processor API changes

- HBASE-16769 - Deprecated PB references from MasterObserver and RegionServerObserver.
- HBASE-17312 - [JDK8] Use default method for Observer Coprocessors. The interface classes of BaseMasterAndRegionObserver, BaseMasterObserver, BaseRegionObserver, BaseRegionServerObserver and BaseWALObserver uses JDK8's 'default' keyword to provide empty and no-op implementations.
- Interface HTableInterface introduces following changes to the methods listed below:

[#] interface CoprocessorEnvironment changes

Change	Result
Abstract method getTable (TableName) has been removed.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getTable (TableName, ExecutorService) has been removed.	A client program may be interrupted by NoSuchMethodError exception.

- Public Audience

The following tables describes the coprocessor changes.

[#] class CoprocessorRpcChannel (1)

Change	Result
This class has become interface.	A client program may be interrupted by IncompatibleClassChangeError or InstantiationException exception depending on the usage of this class.

Class CoprocessorHost<E> changes

Classes that were Audience Private but were removed.

Change	Result
Type of field coprocessors has been changed from java.util.SortedSet<E> to org.apache.hadoop.hbase.util.SortedList<E>.	A client program may be interrupted by NoSuchFieldError exception.

MasterObserver changes

The following changes are introduced to the MasterObserver interface.

[#] interface MasterObserver (14)

Change	Result
Abstract method voidpostCloneSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostCreateTable (ObserverContext<MasterCoproprocessorEnvironment>, HTableDescriptor, HRegionInfo[]) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpostDeleteSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpostGetTableDescriptors (ObserverContext<MasterCoproprocessorEnvironment>, List<HTableDescriptor>) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpostModifyTable (ObserverContext<MasterCoproprocessorEnvironment>, TableName, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpostRestoreSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpostSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreCloneSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreCreateTable (ObserverContext<MasterCoproprocessorEnvironment>, HTableDescriptor, HRegionInfo[]) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreDeleteSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreGetTableDescriptors (ObserverContext<MasterCoproprocessorEnvironment>, List<TableName>, List<HTableDescriptor>) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreModifyTable (ObserverContext<MasterCoproprocessorEnvironment>, TableName, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreRestoreSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.
Abstract method voidpreSnapshot (ObserverContext<MasterCoproprocessorEnvironment>, HBaseProtos.SnapshotDescription, HTableDescriptor) has been removed from this interface.	A client program may be interrupted by NoSuchMethodErrorexception.

RegionObserver interface changes

The following changes are introduced to the RegionObserver interface.

[#] interface RegionObserver (13)

Change	Result
Abstract method voidpostCloseRegionOperation (ObserverContext<RegionCoproprocessorEnvironment>, HRegion.Operation) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostCompactSelection (ObserverContext<RegionCoproprocessorEnvironment>, Store, ImmutableList<StoreFile>) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostCompactSelection (ObserverContext<RegionCoproprocessorEnvironment>, Store, ImmutableList<StoreFile>, CompactionRequest) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostGetClosestRowBefore (ObserverContext<RegionCoproprocessorEnvironment>, byte[], byte[], Result) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method DeleteTrackerpostInstantiateDeleteTracker (ObserverContext<RegionCoproprocessorEnvironment>, DeleteTracker) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostSplit (ObserverContext<RegionCoproprocessorEnvironment>, HRegion, HRegion) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostStartRegionOperation (ObserverContext<RegionCoproprocessorEnvironment>, HRegion.Operation) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method StoreFile.ReaderpostStoreFileReaderOpen (ObserverContext<RegionCoproprocessorEnvironment>, FileSystem, Path, FSDataInputStreamWrapper, long, CacheConfig, Reference, StoreFile.Reader) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpostWALRestore (ObserverContext<RegionCoproprocessorEnvironment>, HRegionInfo, HLogKey, WALEdit) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method InternalScannerpreFlushScannerOpen (ObserverContext<RegionCoproprocessorEnvironment>, Store, KeyValueScanner, InternalScanner) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpreGetClosestRowBefore (ObserverContext<RegionCoproprocessorEnvironment>, byte[], byte[], Result) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method StoreFile.ReaderpreStoreFileReaderOpen (ObserverContext<RegionCoproprocessorEnvironment>, FileSystem, Path, FSDataInputStreamWrapper, long, CacheConfig, Reference, StoreFile.Reader) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method voidpreWALRestore (ObserverContext<RegionCoproprocessorEnvironment>, HRegionInfo, HLogKey, WALEdit) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

WALObserver interface changes

The following changes are introduced to the WALObserver interface.

[#] interface WALObserver

Change	Result
--------	--------

Abstract method void postWALWrite (ObserverContext<WALCoprocessorEnvironment>, HRegionInfo, HLogKey, WALEdit) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method boolean preWALWrite (ObserverContext<WALCoprocessorEnvironment>, HRegionInfo, HLogKey, WALEdit) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

Scheduler changes

Following methods are now changed to abstract.

[#]class RpcScheduler (1)

Change	Result
Abstract method void dispatch (CallRunner) has been removed from this class.	A client program may be interrupted by NoSuchMethodError exception.

hbase-server-1.0.0.jar, RpcScheduler.class package org.apache.hadoop.hbase.ipc

[#] RpcScheduler.dispatch (CallRunner p1) [abstract] : void 1

org/apache/hadoop/hbase/ipc/RpcScheduler.dispatch:(Lorg/apache/hadoop/hbase/ipc/CallRunner;)V

Change	Result
Return value type has been changed from void to boolean.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

The following abstract methods have been removed.

[#]interface PriorityFunction (2)

Change	Result
Abstract method long getDeadline (RPCProtos.RequestHeader, Message) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method int getPriority (RPCProtos.RequestHeader, Message) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

Server API changes

[#] class RpcServer (12)

Change	Result
Type of field CurCall has been changed from java.lang.ThreadLocal<RpcServer.Call> to java.lang.ThreadLocal<RpcCall>.	A client program may be interrupted by NoSuchFieldError exception.
This class became abstract.	A client program may be interrupted by InstantiationException exception.
Abstract method int getNumOpenConnections () has been added to this class.	This class became abstract and a client program may be interrupted by InstantiationException exception.
Field callQueueSize of type org.apache.hadoop.hbase.util.Counter has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field connectionList of type java.util.List<RpcServer.Connection> has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field maxIdleTime of type int has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field numConnections of type int has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field port of type int has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.

Field purgeTimeout of type long has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field responder of type RpcServer.Responder has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field socketSendBufferSize of type int has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field thresholdIdleConnections of type int has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.

Following abstract method has been removed:

Change	Result
Abstract method Pair<Message,CellScanner>call (BlockingService, Descriptors.MethodDescriptor, Message, CellScanner, long, MonitoredRPCHandler) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

Replication and WAL changes

HBASE-18733: WALKey has been purged completely. Following are the changes to the WALKey:

[#] classWALKey (8)

Change	Result
Access level of field clusterIds has been changed from protected to private.	A client program may be interrupted by IllegalAccessError exception.
Access level of field compressionContext has been changed from protected to private.	A client program may be interrupted by IllegalAccessError exception.
Access level of field encodedRegionName has been changed from protected to private.	A client program may be interrupted by IllegalAccessError exception.
Access level of field tablename has been changed from protected to private.	A client program may be interrupted by IllegalAccessError exception.
Access level of field writeTime has been changed from protected to private.	A client program may be interrupted by IllegalAccessError exception.

Following fields have been removed:

Change	Result
Field LOG of type org.apache.commons.logging.Log has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field VERSION of type WALKey.Version has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field logSeqNum of type long has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.

Following are the changes to the WALEdit.class: hbase-server-1.0.0.jar, WALEdit.class package org.apache.hadoop.hbase.regionserver.wal

Deprecated APIs or coprocessor

HBASE-16769 - Protocol Buffers (PB) references from MasterObserver and RegionServerObserver has been removed.

Admin Interface API changes

You cannot administer a CDP Runtime Data Hub cluster using a client that includes RelocationAdmin, ACC, Thrift and REST usage of Admin ops. Methods returning protobufs have been changed to return POJOs instead. Returns have changed from void to Future for async methods. HBASE-18106 - Admin.listProcedures and

Admin.listLocks were renamed to getProcedures and getLocks. MapReduce makes use of Admin doing following admin.getClusterStatus() to calculate Splits.

- Thrift usage of Admin API:

```
compact(ByteBuffer) createTable(ByteBuffer, List<ColumnDescriptor>) deleteTable(ByteBuffer) disableTable(ByteBuffer) enableTable(ByteBuffer) getTableNames() majorCompact(ByteBuffer)
```

- REST usage of Admin API:

```
hbase-rest org.apache.hadoop.hbase.rest RootResource getTableList() TableName[] tableNames = servlet.getAdmin().listTableNames(); SchemaResource delete(UriInfo) Admin admin = servlet.getAdmin(); update(TableSchemaModel, boolean, UriInfo) Admin admin = servlet.getAdmin(); StorageClusterStatusResource get(UriInfo) ClusterStatus status = servlet.getAdmin().getClusterStatus(); StorageClusterVersionResource get(UriInfo) model.setVersion(servlet.getAdmin().getClusterStatus().getHBaseVersion()); TableResource exists() return servlet.getAdmin().tableExists(TableName.valueOf(table));
```

[#] interface Admin (9)

Following are the changes to the Admin interface:

Change	Result
Abstract method createTableAsync (HTableDescriptor, byte[][]) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method disableTableAsync (TableName) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method enableTableAsync (TableName) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getCompactionState (TableName) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getCompactionStateForRegion (byte[]) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method isSnapshotFinished (HBaseProtos.SnapshotDescription) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method snapshot (String, TableName, HBaseProtos.SnapshotDescription.Type) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method snapshot (HBaseProtos.SnapshotDescription) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method takeSnapshotAsync (HBaseProtos.SnapshotDescription) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

[#] Admin.createTableAsync (HTableDescriptor p1, byte[][] p2) [abstract] : void 1

org/apache/hadoop/hbase/client/Admin.createTableAsync:(Lorg/apache/hadoop/hbase/HTableDescriptor;[[B)V

Change	Result
Return value type has been changed from void to java.util.concurrent.Future<java.lang.Void>.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

[#] Admin.disableTableAsync (TableName p1) [abstract] : void 1

org/apache/hadoop/hbase/client/Admin.disableTableAsync:(Lorg/apache/hadoop/hbase/TableName;)V

Change	Result
Return value type has been changed from void to java.util.concurrent.Future<java.lang.Void>.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

Admin.enableTableAsync (TableName p1) [abstract] : void 1

org/apache/hadoop/hbase/client/Admin.enableTableAsync:(Lorg/apache/hadoop/hbase/TableName;)V

Change	Result
Return value type has been changed from void to java.util.concurrent.Future<java.lang.Void>.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

Admin.enableTableAsync (TableName p1) [abstract] : void 1

org/apache/hadoop/hbase/client/Admin.getCompactionState:(Lorg/apache/hadoop/hbase/TableName;)Lorg/apache/hadoop/hbase/protobuf/generated/AdminProtos\$GetRegionInfoResponse\$CompactionState;

Change	Result
Return value type has been changed from org.apache.hadoop.hbase.protobuf.generated.AdminProtos.GetRegionInfoResponse.CompactionState to CompactionState.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

[#] Admin.getCompactionStateForRegion (byte[] p1) [abstract] :
AdminProtos.GetRegionInfoResponse.CompactionState 1

org/apache/hadoop/hbase/client/Admin.getCompactionStateForRegion:([B)Lorg/apache/hadoop/hbase/protobuf/generated/AdminProtos\$GetRegionInfoResponse\$CompactionState;

Change	Result
Return value type has been changed from org.apache.hadoop.hbase.protobuf.generated.AdminProtos.GetRegionInfoResponse.CompactionState to CompactionState.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

HTableDescriptor and HColumnDescriptor changes

HTableDescriptor and HColumnDescriptor has become interfaces and you can create it through Builders. HCD has become CFD. It no longer implements writable interface. package org.apache.hadoop.hbase[#] class HColumnDescriptor (1)

Change	Result
Removed super-interface org.apache.hadoop.io.WritableComparable<HColumnDescriptor>.	A client program may be interrupted by NoSuchMethodError exception.

HColumnDescriptor in 1.0.0 {code} @InterfaceAudience.Public @InterfaceStability.Evolving public class HColumnDescriptor implements WritableComparable<HColumnDescriptor> { {code}

HColumnDescriptor in 2.0 {code} @InterfaceAudience.Public @Deprecated // remove it in 3.0 public class HColumnDescriptor implements ColumnFamilyDescriptor, Comparable<HColumnDescriptor> { {code}

For META_TABLEDESC, the maker method had been deprecated already in HTD in 1.0.0. OWNER_KEY is still in HTD.

class HTableDescriptor (3)

hbase-client-1.0.0.jar, HTableDescriptor.class package org.apache.hadoop.hbase

Change	Result
--------	--------

Removed super-interface org.apache.hadoop.io.WritableComparable<HTableDescriptor>.	A client program may be interrupted by NoSuchMethodError exception.
Field META_TABLEDESC of type HTableDescriptor has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.

[#] HTableDescriptor.getColumnFamilies () : HColumnDescriptor[] (1)

org/apache/hadoop/hbase/HTableDescriptor.getColumnFamilies():[Lorg/apache/hadoop/hbase/HColumnDescriptor;[#] class HColumnDescriptor (1)

Change	Result
Return value type has been changed from HColumnDescriptor[] to client.ColumnFamilyDescriptor[].	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

[#] interface Table (4)

Change	Result
Abstract method batch (List<?>) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method batchCallback (List<?>, Batch.Callback<R>) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getWriteBufferSize () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method setWriteBufferSize (long) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

Deprecated buffer methods

- HBASE-13298- Clarify if Table.{set|get}WriteBufferSize() is deprecated or not.
- LockTimeoutException and OperationConflictException classes have been removed.

class OperationConflictException (1)

Change	Result
This class has been removed.	A client program may be interrupted by NoClassDefFoundError exception.

class class LockTimeoutException (1)

Change Result This class has been removed. A client program may be interrupted by NoClassDefFoundError exception.

Filter API changes

Following methods have been removed: package org.apache.hadoop.hbase.filter

[#] class Filter (2)

Change	Result
Abstract method getNextKeyHint (KeyValue) has been removed from this class.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method transform (KeyValue) has been removed from this class.	A client program may be interrupted by NoSuchMethodError exception.

- HBASE-12296 Filters should work with ByteBufferedCell.
- HConnection is removed in HBase 2.0.
- RegionLoad and ServerLoad internally moved to shaded PB.

[#] class RegionLoad (1)

Change	Result
Type of field regionLoadPB has been changed from <code>protobuf.generated.ClusterStatusProtos.RegionLoad</code> to <code>shaded.protobuf.generated.ClusterStatusProtos.RegionLoad</code> .	A client program may be interrupted by <code>NoSuchFieldError</code> exception.

HBASE-15783:AccessControlConstants#OP_ATTRIBUTE_ACL_STRATEGY_CELL_FIRST is not used any more.
package `org.apache.hadoop.hbase.security.access`

[#] interface AccessControlConstants (3)

Change	Result
Field <code>OP_ATTRIBUTE_ACL_STRATEGY</code> of type <code>java.lang.String</code> has been removed from this interface.	A client program may be interrupted by <code>NoSuchFieldError</code> exception.
Field <code>OP_ATTRIBUTE_ACL_STRATEGY_CELL_FIRST</code> of type <code>byte[]</code> has been removed from this interface.	A client program may be interrupted by <code>NoSuchFieldError</code> exception.
Field <code>OP_ATTRIBUTE_ACL_STRATEGY_DEFAULT</code> of type <code>byte[]</code> has been removed from this interface.	A client program may be interrupted by <code>NoSuchFieldError</code> exception.

`ServerLoad` returns long instead of int 1

`hbase-client-1.0.0.jar`, `ServerLoad.class` package `org.apache.hadoop.hbase`

[#] `ServerLoad.getNumberOfRequests () : int 1`

`org/apache/hadoop/hbase/ServerLoad.getNumberOfRequests():I`

Change	Result
Return value type has been changed from int to long.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by <code>NoSuchMethodError</code> exception.

[#] `ServerLoad.getNumberOfRequests () : int 1`

`org/apache/hadoop/hbase/ServerLoad.getReadRequestsCount():I`

Change	Result
Return value type has been changed from int to long.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by <code>NoSuchMethodError</code> exception.

b[#] `ServerLoad.getTotalNumberOfRequests () : int 1`

`org/apache/hadoop/hbase/ServerLoad.getTotalNumberOfRequests():I`

Change	Result
Return value type has been changed from int to long.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by <code>NoSuchMethodError</code> exception.

[#] `ServerLoad.getWriteRequestsCount () : int 1`

`org/apache/hadoop/hbase/ServerLoad.getWriteRequestsCount():I`

Change	Result
Return value type has been changed from int to long.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by <code>NoSuchMethodError</code> exception.

- HBASE-13636 Remove deprecation for HBASE-4072 (Reading of `zoo.cfg`)

- HConstants are removed. HBASE-16040 Remove configuration "hbase.replication"

[#]class HConstants (6)

Change	Result
Field DEFAULT_HBASE_CONFIG_READ_ZOOKEEPER_CONFIG of type boolean has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field HBASE_CONFIG_READ_ZOOKEEPER_CONFIG of type java.lang.String has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field REPLICATION_ENABLE_DEFAULT of type boolean has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field REPLICATION_ENABLE_KEY of type java.lang.String has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field ZOOKEEPER_CONFIG_NAME of type java.lang.String has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.
Field ZOOKEEPER_USEMULTI of type java.lang.String has been removed from this class.	A client program may be interrupted by NoSuchFieldError exception.

HBASE-18732: [compat 1-2] HBASE-14047 removed Cell methods without deprecation cycle.

[#]interface Cell 5

Change	Result
Abstract method getFamily () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getMvccVersion () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getQualifier () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getRow () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.
Abstract method getValue () has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

HBASE-18795:Expose KeyValue.getBuffer() for tests alone. Allows KV#getBuffer in tests only that was deprecated previously.

Region scanner changes

[#]interface RegionScanner (1)

Change	Result
Abstract method boolean nextRow (List<Cell>, int) has been removed from this interface.	A client program may be interrupted by NoSuchMethodError exception.

StoreFile changes

[#] class StoreFile (1)

Change	Result
This class became interface.	A client program may be interrupted by IncompatibleClassChangeError or InstantiationException exception dependent on the usage of this class.

Mapreduce changes

HFile*Format has been removed.

ClusterStatus changes

HBASE-15843: Replace RegionState.getRegionInTransition() Map with a Set hbase-client-1.0.0.jar, ClusterStatus.class package org.apache.hadoop.hbase

[#] ClusterStatus.getRegionsInTransition () : Map<String,RegionState> 1

org/apache/hadoop/hbase/ClusterStatus.getRegionsInTransition:()Ljava/util/Map;

Change	Result
Return value type has been changed from java.util.Map<java.lang.String, master.RegionState> to java.util.List<master.RegionState>.	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

Other changes in ClusterStatus include removal of convert methods that were no longer necessary after purge of PB from API.

Purge of Protocol Buffers from API

Protocol Buffers (PB) has been deprecated in APIs.

[#] HBaseSnapshotException.getSnapshotDescription () : HBaseProtos.SnapshotDescription 1

```
org/apache/hadoop/hbase/snapshot/HBaseSnapshotException.getSnapshotDescription:()Lorg/apache/hadoop/hbase/protobuf/generated/HBaseProtos$SnapshotDescription;
```

Change	Result
Return value type has been changed from org.apache.hadoop.hbase.protobuf.generated.HBaseProtos.SnapshotDescription to org.apache.hadoop.hbase.client.SnapshotDescription	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

HBASE-15609: Remove PB references from Result, DoubleColumnInterpreter and any such public facing class for 2.0. hbase-client-1.0.0.jar, Result.class package org.apache.hadoop.hbase.client

[#] Result.getStats () : ClientProtos.RegionLoadStats 1

org/apache/hadoop/hbase/client/Result.getStats:()Lorg/apache/hadoop/hbase/protobuf/generated/ClientProtos\$RegionLoadStats;

Change	Result
Return value type has been changed from org.apache.hadoop.hbase.protobuf.generated.ClientProtos.RegionLoadStats to RegionLoadStats	This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchMethodError exception.

PrettyPrinter changes

hbase-server-1.0.0.jar, HFilePrettyPrinter.class package org.apache.hadoop.hbase.io.hfile

Change	Result
--------	--------

Return value type has been changed from void to int.

This method has been removed because the return type is part of the method signature. A client program may be interrupted by NoSuchElementException exception.

Technical Service Bulletins

TSB 2021-494: Accumulated WAL Files Cannot be Cleaned up When Using Phoenix Secondary Global Indexes

The Write-ahead-log (WAL) files for Phoenix tables that have secondary global indexes defined on them, cannot be automatically cleaned up by HBase, leading to excess storage usage and possible error due to filling up the storage. Accumulated WAL files can lead to lengthy restart times as they must all be played back to ensure no dataloss occurs on restart. This can have follow-on HDFS impact if the number of WAL files overwhelm HDFS Name Node.

Upstream JIRA

- [HBASE-20781](#)
- [HBASE-25459](#)
- [PHOENIX-5250](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-494: Accumulated WAL Files Cannot be Cleaned up When Using Phoenix Secondary Global Indexes](#)

TSB 2021-506: Active HBase MOB files can be removed

Actively used MOB files can be deleted by MobFileCleanerChore due to incorrect serialization of reference file names. This is causing data loss on MOB-enabled tables.

Upstream JIRA

- [HBASE-23723](#)
- [HBASE-25970](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-506: Active HBase MOB files can be removed](#)

TSB 2022-569: HBase normalizer can cause table inconsistencies by merging non-adjacent regions

The normalizer in HBase is a background job responsible for splitting or merging HBase regions to optimize the number of regions and the distribution of the size of the regions in HBase tables. Due to the bug described in HBASE-24376, the normalizer can cause region inconsistencies (region overlaps/holes) by merging non-adjacent regions.

Upstream JIRA

[HBASE-24376](#)

Knowledge article

For the latest update on this issue, see the corresponding Knowledge article: [TSB 2022-569: HBase normalizer can cause table inconsistencies by merging non-adjacent regions](#)

Known Issues in HDFS

This topic describes known issues and unsupported features for using HDFS in this release of Cloudera Runtime.

CDPD-2946: Slow reading and writing of erasure-coded files

The ISA-L library is not packaged with HDFS as a result of which HDFS erasure coding falls back to the Java implementation which is much slower than the native Hadoop implementation. This slows down the reading and writing of erasure-coded files.

Unsupported Features

The following HDFS features are currently not supported in Cloudera Data Platform:

- ACLs for the NFS gateway ([HADOOP-11004](#))
- Aliyun Cloud Connector ([HADOOP-12756](#))
- Allow HDFS block replicas to be provided by an external storage system ([HDFS-9806](#))
- Consistent standby Serving reads ([HDFS-12943](#))
- Cost-Based RPC FairCallQueue ([HDFS-14403](#))
- HDFS Router Based Federation ([HDFS-10467](#))
- More than two NameNodes ([HDFS-6440](#))
- NameNode Federation ([HDFS-1052](#))
- NameNode Port-based Selective Encryption ([HDFS-13541](#))
- Non-Volatile Storage Class Memory (SCM) in HDFS Cache Directives ([HDFS-13762](#))
- OpenStack Swift ([HADOOP-8545](#))
- SFTP FileSystem ([HADOOP-5732](#))
- Storage policy satisfier ([HDFS-10285](#))
- Transparent Data Encryption ([HDFS-6134](#))
- Upgrade Domain ([HDFS-7541](#))

Technical Service Bulletins

TSB 2021-406: CVE-2020-9492 Hadoop filesystem bindings (ie: webhdfs) allows credential stealing

WebHDFS clients might send SPNEGO authorization header to remote URL without proper verification. A maliciously crafted request can trigger services to send server credentials to a webhdfs path (ie: webhdfs://...) for capturing the service principal.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB-2021 406: CVE-2020-9492 Hadoop filesystem bindings \(ie: webhdfs\) allows credential stealing](#)

Known Issues in Apache Hive

This topic describes known issues and workarounds for using Hive in this release of Cloudera Runtime.

OPSAPS-54299 Installing Hive on Tez and HMS in the incorrect order causes HiveServer failure

You need to install Hive on Tez and HMS in the correct order; otherwise, HiveServer fails. You need to install additional HiveServer roles to Hive on Tez, not the Hive service; otherwise, HiveServer fails.

Follow instructions on [Installing Hive on Tez](#).

CDPD-23041: DROP TABLE on a table having an index does not work

If you migrate a Hive table to CDP having an index, DROP TABLE does not drop the table. Hive no longer supports indexes ([HIVE-18448](#)). A foreign key constraint on the indexed table prevents dropping the table. Attempting to drop such a table results in the following error:

```
java.sql.BatchUpdateException: Cannot delete or update a parent
row: a foreign key constraint fails ("hive"."IDX$", CONSTRAINT "
IDX$_FK1" FOREIGN KEY ("ORIG_TBL_ID") REFERENCES "TBLS ("TBL_ID"
))
```

There are two workarounds:

- Drop the foreign key "IDX\$_FK1" on the "IDX\$" table within the metastore. You can also manually drop indexes, but do not cascade any drops because the IDX\$ table includes references to "TBLS".
- Launch an older version of Hive, such as Hive 2.3 that includes IDX\$ in the DDL, and then drop the indexes as described in [Language Manual Indexing](#).

Apache Issue: [Hive-24815](#)

Technical Service Bulletins

TSB 2021-480/1: Hive produces incorrect query results when skipping a header in a binary file

In CDP, setting the table property `skip.header.line.count` to greater than 0 in a table stored in a binary format, such as Parquet, can cause incorrect query results. The skip header property is intended for use with Text files and typically used with CSV files. The issue is not present when you run the query on a Text file that sets the skip header property to 1 or greater.

Upstream JIRA

[Apache Jira: HIVE-24827](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-480.1: Hive produces incorrect query results when skipping a header in a binary file](#)

TSB 2021-480/2: Hive ignores the property to skip a header or footer in a compressed file

In CDP, setting the table properties `skip.header.line.count` and `skip.footer.line.count` to greater than 0 in a table stored in a compressed format, such as bzip2, can cause incorrect results from `SELECT *` or `SELECT COUNT (*)` queries.

Upstream JIRA

[Apache Jira: HIVE-24224](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-480.2: Hive ignores the property to skip a header or footer in a compressed file](#)

TSB 2021-482: Race condition in subdirectory delete/rename causes hive jobs to fail

Multiple threads try to perform a rename operation on s3. One of the threads fails to perform a rename operation, causing an error. Hive logs will report "HiveException: Error moving ..." and the log will contain an error line starting with "Exception when loading partition " -all paths listed with s3a:// prefixes.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-482: Race condition in subdirectory delete/rename causes Hive jobs to fail](#)

TSB 2021-501: JOIN queries return wrong result for join keys with large size in Hive

JOIN queries return wrong results when performing joins on large size keys (larger than 255 bytes). This happens when the fast hash table join algorithm is enabled, which is enabled by default.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-501: JOIN queries return wrong result for join keys with large size in Hive](#)

TSB 2021-518: Incorrect results returned when joining two tables with different bucketing versions

Incorrect results are returned when joining two tables with different bucketing versions, and with the following Hive configurations: `set hive.auto.convert.join = false` and `set mapreduce.job.reduces = any custom value`.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-518: Incorrect results returned when joining two tables with different bucketing versions](#)

TSB 2023-627: IN/OR predicate on binary column returns wrong result

An IN or an OR predicate involving a binary datatype column may produce wrong results. The OR predicate is converted to an IN due to the setting `hive.optimize.point.lookup` which is true by default. Only binary data types are affected by this issue. See <https://issues.apache.org/jira/browse/HIVE-26235> for example queries which may be affected.

Upstream JIRA[HIVE-26235](#)**Knowledge article**

For the latest update on this issue, see the corresponding Knowledge article: [TSB 2023-627: IN/OR predicate on binary column returns wrong result](#)

Known Issues in Hue

This topic describes known issues and workarounds for using Hue in this release of Cloudera Runtime.

CDPD-18959: Due to tablib upgrade for Hue in CDP from 0.10.x to 0.14.x release which is generally used for Python3 release, the CSV download has a special character in the dataset which fails with the error message "UnicodeDecodeError: 'ascii' codec can't decode byte 0xef in position 8: ordinal not in range(128)"

Workaround:

- `cd /opt/cloudera/parcels/CDH/lib/`
- Backup the hue folder `cp -R hue hue_orgi`
- `cd hue`
- Install python packages via `pippip install wheel`
- Install python packages for Hue
 - `./build/env/bin/pip install wheel.`
 - `/build/env/bin/pip setuptools==44.1.0`
 - `./build/env/bin/pip install tablib==0.12.1`
- Restart Hue and check again

Hive and Impala query editors do not work with TLS 1.2

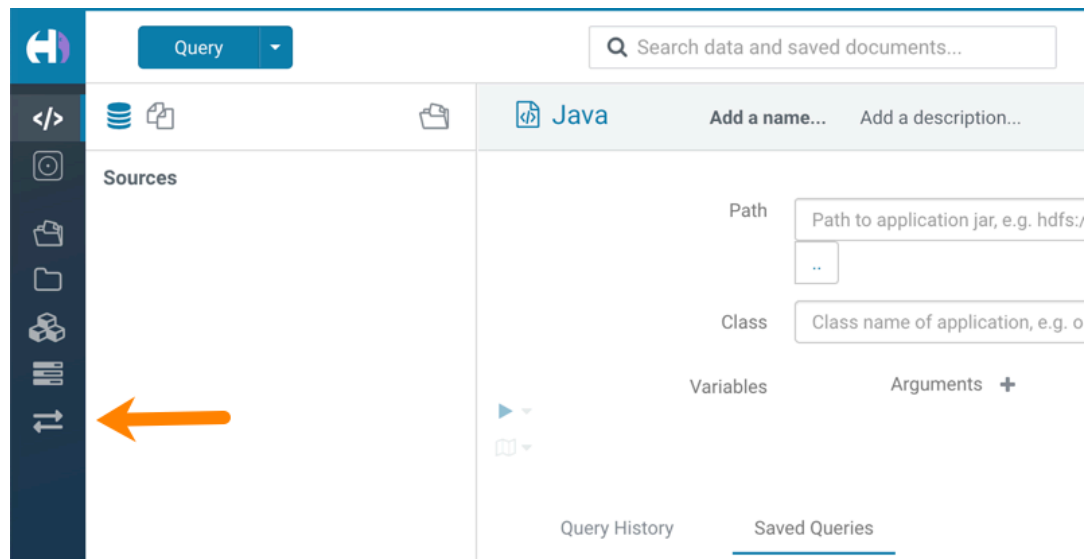
Problem: If Hive or Impala engines are using TLS version 1.2 on your CDP cluster, then you won't be able to run queries from the Hue Hive or Impala query editor.

Workaround: You must apply the following patch: HUE-9508. Contact Cloudera Support for help on applying the software patch.

Hue Importer is not supported in the Data Engineering template

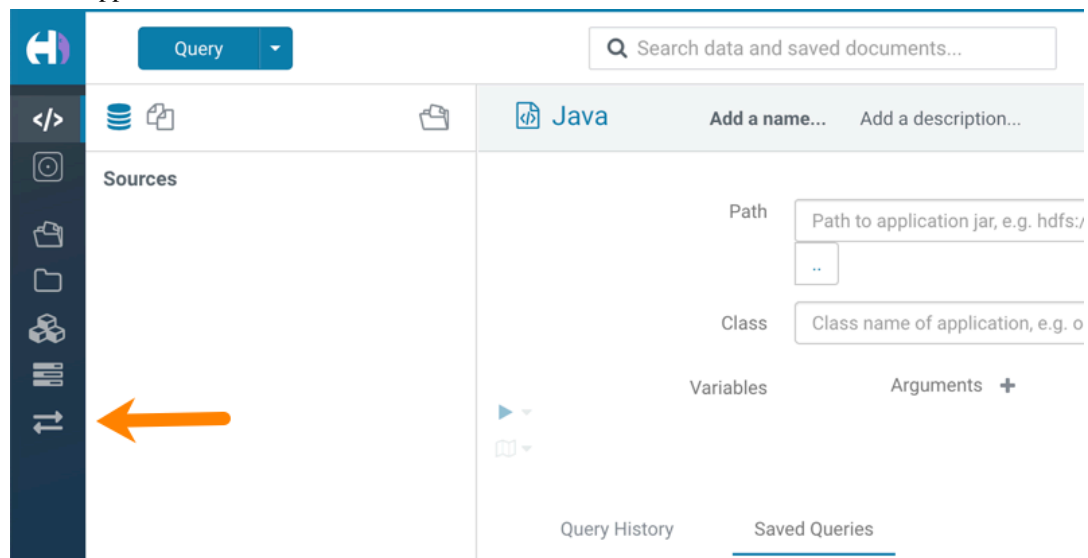
When you create a Data Hub cluster using the Data Engineering template, the Importer application is not supported in Hue.

Figure 1: Hue web UI showing Importer icon on the left assist panel



Hue Importer is not supported in the Data Engineering template

When you create a Data Hub cluster using the Data Engineering template, the Importer application is not supported in Hue:



CDPD-3501: Hue-Atlas configuration information is missing on Data Mart clusters.

Problem: The configuration file `hive-conf%2Fatlas-application.properties` is missing on Data Mart clusters because Apache Hive is not installed. This properties file is needed for the Hue integration with Apache Atlas.

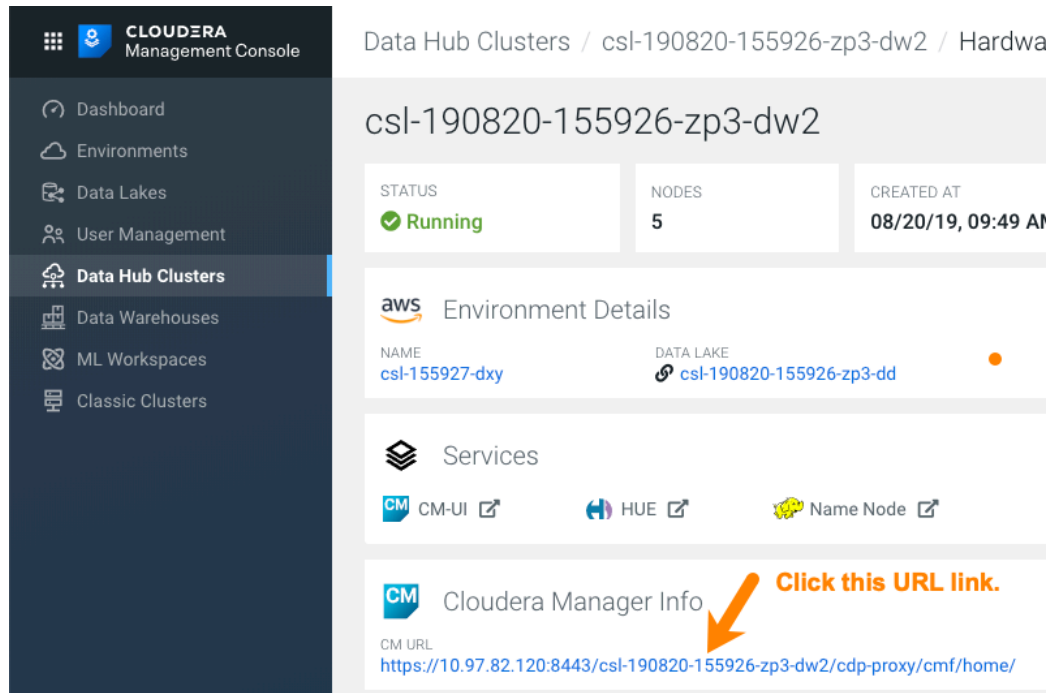
Workaround:



Note: To make the following configuration change, you must have administrative permissions on the Data Mart cluster.

1. Log in to the CDP web interface and navigate to the Data Hub service.
2. On the Data Hub Clusters page, click the Data Mart cluster you want to work on.

3. On the Data Mart cluster page, click the URL link to Cloudera Manager Info:



Data Hub Clusters / csl-190820-155926-zp3-dw2 / Hardware

csl-190820-155926-zp3-dw2

STATUS: ✔ Running | NODES: 5 | CREATED AT: 08/20/19, 09:49 AM

Environment Details

NAME: csl-155927-dxy | DATA LAKE: csl-190820-155926-zp3-dd

Services

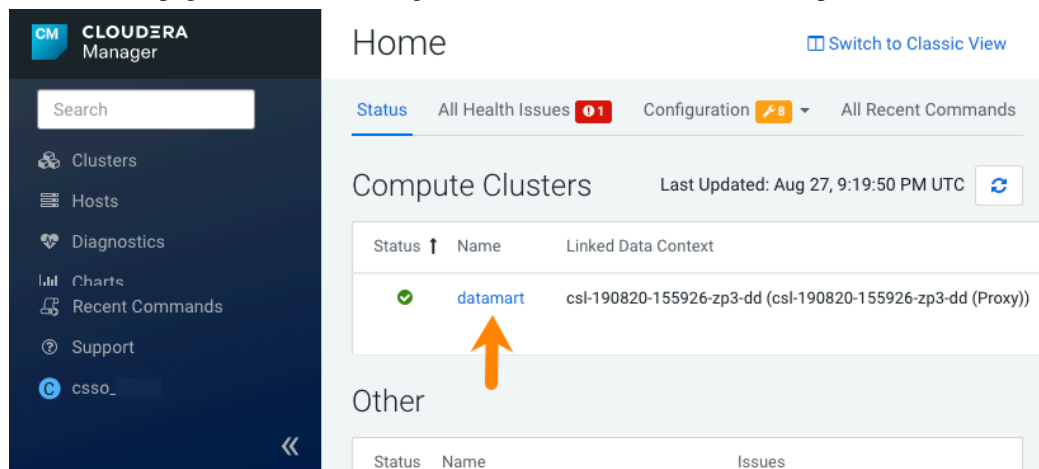
CM-UI | HUE | Name Node

Cloudera Manager Info

CM URL: <https://10.97.82.120:8443/csl-190820-155926-zp3-dw2/cdp-proxy/cm/home/>

Click this URL link.

4. On the Home page of Cloudera Manager, click the cluster name under Compute Clusters:



Home [Switch to Classic View](#)

Status | All Health Issues 1 | Configuration 8 | All Recent Commands

Compute Clusters Last Updated: Aug 27, 9:19:50 PM UTC [Refresh](#)

Status	Name	Linked Data Context
✔	datamart	csl-190820-155926-zp3-dd (csl-190820-155926-zp3-dd (Proxy))

Other

Status	Name	Issues
--------	------	--------

5. In the cluster page in the Status column under Compute Cluster, Cloudera Runtime, click the link to Hue:

The screenshot shows the Cloudera Manager interface. On the left is a dark sidebar with the 'CLUSTERA Manager' logo and a search bar. Below the search bar are links for Clusters, Hosts, Diagnostics, and Charts. At the bottom of the sidebar are links for Recent Commands, Support, and CSO. The main content area is titled 'datamart' and has tabs for Status, Health Issues, and Configuration (with a '2' icon). The 'Status' tab is active, showing a list of services: zookeeper, solr, kafka, ranger, atlas, Knox, and hive. Below this list is a section for 'Compute Cluster, Cloudera Runtime 7.0.0 (Parcels)' which lists services: 5 Hosts, HDFS, Hue, Impala, Knox, and csl-190820-15... An orange arrow points to the 'Hue' service link.

6. On the Hue page, click the Configuration tab to view the configuration properties for Hue.
7. In the search text box, type safety and press Enter to locate the Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini, which appears at the top of the configuration parameters list.
8. Append the following configuration information to the existing configuration information in the Safety Valve and click Save Changes:

```
[metadata]
[[catalog]]
interface=atlas
api_url=http://master0.cloudera.site:21000/api/atlas/
kerberos_enabled=true
```

9. Restart the Hue service for the configuration change to take effect.

BalancerMember worker hostname too long error

You may see the following error message while starting the Hue Load Balancer:

```
BalancerMember worker hostname (xxx-xxxxxxx-xxxxxxxxxx-xxxxxxx
.xxxxxx-xxxxxx-xxxxxx.example.site) too long.
```

Cloudera Manager displays this error when you create a Data Hub cluster using the Data Engineering template and the Hue Load Balancer worker node name has exceeded 64 characters. In a CDP Public Cloud deployment, the system automatically generates the Load Balancer worker node name through AWS or Azure.

For example, if you specify `cdp-123456-scalecluster` as the cluster name, CDP creates `cdp-123456-scalecluster-master2.repro-aw.a123-4a5b.example.site` as the worker node name.

Workaround: Specify a shorter cluster name while creating a Data Hub cluster so that the final worker node name does not cross 64 characters.

For example, `cdp-123456-scale`.

Unsupported features

Importing and exporting Oozie workflows across clusters and between different CDH versions is not supported

You can export Oozie workflows, schedules, and bundles from Hue and import them only within the same cluster if the cluster is unchanged. You can migrate bundle and coordinator jobs with their workflows only if their arguments have not changed between the old and the new cluster. For example, hostnames, NameNode, Resource Manager names, YARN queue names, and all the other parameters defined in the `workflow.xml` and `job.properties` files.

Using the import-export feature to migrate data between clusters is not recommended. To migrate data between different versions of CDH, for example, from CDH 5 to CDP 7, you must take the dump of the Hue database on the old cluster, restore it on the new cluster, and set up the database in the new environment. Also, the authentication method on the old and the new cluster should be the same because the Oozie workflows are tied to a user ID, and the exact user ID needs to be present in the new environment so that when a user logs into Hue, they can access their respective workflows.



Note: Migrating Oozie workflows from HDP clusters is not supported.

PySpark and SparkSQL are not supported with Livy in Hue

Hue does not support configuring and using PySpark and SparkSQL with Livy in CDP Private Cloud Base.

Technical Service Bulletins

TSB 2021-487: Cloudera Hue is vulnerable to Cross-Site Scripting attacks

Multiple Cross-Site Scripting (XSS) vulnerabilities of Cloudera Hue have been found. They allow JavaScript code injection and execution in the application context.

- CVE-2021-29994 - The Add Description field in the Table schema browser does not sanitize user inputs as expected.
- CVE-2021-32480 - Default Home direct button in Filebrowser is also susceptible to XSS attack.
- CVE-2021-32481 - The Error snippet dialog of the Hue UI does not sanitize user inputs.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-487: Cloudera Hue is vulnerable to Cross-Site Scripting attacks \(CVE-2021-29994, CVE-2021-32480, CVE-2021-32481\)](#)

Known Issues in Apache Impala

This topic describes known issues and workarounds for using Impala in this release of Cloudera Runtime.
Queries stuck on failed HDFS calls and not timing out

In Impala 3.2 and higher, if the following error appears multiple times in a short duration while running a query, it would mean that the connection between the impalad and the HDFS NameNode is in a bad state.

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed
to finish before the <hdfs_operation_timeout_sec> second timeout
"
```

In Impala 3.1 and lower, the same issue would cause Impala to wait for a long time or not respond without showing the above error message.

Workaround: Restart the impalad.

Apache JIRA: HADOOP-15720

Impala should tolerate bad locale settings

If the LC_* environment variables specify an unsupported locale, Impala does not start.

Workaround: Add LC_ALL="C" to the environment settings for both the Impala daemon and the Statestore daemon.

Apache JIRA: IMPALA-532

Configuration to prevent crashes caused by thread resource limits

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal
error: Unable to allocate section memory!
terminate called after throwing an instance of 'boost::exception_
detail::clone_impl<boost::exception_detail::error_info_injector<
boost::thread_resource_error> >'
```

Workaround: To prevent such errors, configure each host running an impalad daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

Add the following lines in /etc/security/limits.conf:

```
impala soft nproc 262144
impala hard nproc 262144
```

Apache JIRA: IMPALA-5605

Avro Scanner fails to parse some schemas

The default value in Avro schema must match type of first union type, e.g. if the default value is null, then the first type in the UNION must be "null".

Workaround: Swap the order of the fields in the schema specification. For example, use ["null", "string"] instead of ["string", "null"]. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Apache JIRA: IMPALA-635

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the impalad daemon.

Workaround: To monitor overall memory usage, use the top command, or add the memory figures in the Impala web UI /memz tab to JVM memory usage shown on the /metrics tab.

Apache JIRA: IMPALA-691

Ranger audit logs for applying column masking policies missing

Impala is not producing these logs.

Workaround: None.

Apache JIRA: IMPALA-9350

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Apache JIRA: IMPALA-1024

Incorrect results with basic predicate on CHAR typed column

When comparing a CHAR column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Workaround: Use the RPAD() function to blank-pad literals compared with CHAR columns to the expected length.

Apache JIRA: IMPALA-1652

ImpalaODBC: Can not get the value in the SQLGetData(m-x th column) after the SQLBindCol(m th column)

If the ODBC SQLGetData is called on a series of columns, the function calls must follow the same order as the columns. For example, if data is fetched from column 2 then column 1, the SQLGetData call for column 1 returns NULL.

Workaround: Fetch columns in the same order they are defined in the table.

Apache JIRA: IMPALA-1792

Casting scenarios with invalid/inconsistent results

Using a CAST() function to convert large literal values to smaller types, or to convert special values such as NaN or Inf, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Apache JIRA: IMPALA-1821

A failed CTAS does not drop the table if the insert fails

If a CREATE TABLE AS SELECT operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed CREATE TABLE AS SELECT

Apache JIRA: IMPALA-2005

% escaping does not work correctly when occurs at the end in a LIKE clause

If the final character in the RHS argument of a LIKE operator is an escaped \% character, it does not match a % final character of the LHS argument.

Metadata operations that change the state of a table, like `COMPUTE STATS` or `ALTER RECOVER PARTITIONS`, may delay metadata propagation of unrelated unloaded tables triggered by statements like `DESCRIBE` or `SELECT` queries.

Workaround:

Apache JIRA: [IMPALA-6671](#)

Impala does not support Heimdal Kerberos

Apache JIRA: [IMPALA-7072](#)

Technical Service Bulletins

TSB-2021-485: Impala returns fewer rows from parquet tables on S3

[IMPALA-10310](#) was an issue in Impala's Parquet page filtering code where the scanner did not reset state appropriately when transitioning from the first row group to subsequent row groups in a single split. This caused data from the subsequent row groups to be skipped incorrectly, leading to incorrect query results. This issue cannot occur when the Parquet page filtering is disabled by setting `PARQUET_READ_PAGE_INDEX=false`.

The issue is more likely to be encountered on S3/ADLS/ABFS/etc, because Spark is sometimes configured to write 128MB row groups and the `PARQUET_OBJECT_STORE_SPLIT_SIZE` is 256MB. This makes it more likely for Impala to process two row groups in a single split.

Parquet page filtering only works based on the min/max statistics, therefore the comparison operators it supports are “=”, “<”, “>”, “<=”, and “>=”. These operators are impacted by this bug. Expressions such as “!=”, 'LIKE' or the expressions including UDF do not use parquet page filtering.

The `PARQUET_OBJECT_STORE_SPLIT_SIZE` parameter is introduced in Impala 3.3 by [IMPALA-5843](#). This means that older versions of Impala do not have this issue.

Upstream JIRA

- [IMPALA-5843](#)
- [IMPALA-10310](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-485: Impala returns fewer rows from parquet tables on S3](#)

TSB 2021-502: Impala logs the session / operation secret on most RPCs at INFO level

Impala logs contain the session / operation secret. With this information a person who has access to the Impala logs might be able to hijack other users' sessions. This means the attacker is able to execute statements for which they do not have the necessary privileges otherwise. Impala deployments where Apache Sentry or Apache Ranger authorization is enabled may be vulnerable to privilege escalation. Impala deployments where audit logging is enabled may be vulnerable to incorrect audit logging.

Restricting access to the Impala logs that expose secrets will reduce the risk of an attack. Additionally, restricting access to trusted users for the Impala deployment will also reduce the risk of an attack. Log redaction techniques can be used to redact secrets from the logs. For more information, see the *Cloudera Manager documentation*.

For log redaction, users can create a rule with a search pattern: `secret \((string\) [=].*` And the replacement could be for example: `secret=LOG-REDACTED`

Upstream JIRA

[IMPALA-10600](#)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-502: Impala logs the session / operation secret on most RPCs at INFO level](#)

Known Issues in Apache Kafka

This topic describes known issues and unsupported features for using Kafka in this release of Cloudera Runtime.

Known Issues

Topics created with the kafka-topics tool are only accessible by the user who created them when the deprecated --zookeeper option is used

By default all created topics are secured. However, when topic creation and deletion is done with the kafka-topics tool using the --zookeeper option, the tool talks directly to Zookeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. As a result, if the --zookeeper option is used, only the user who created the topic will be able to carry out administrative actions on it. In this scenario Kafka will not have permissions to perform tasks on topics created this way.

Workaround: Use kafka-topics with the --bootstrap-server option that does not require direct access to Zookeeper.

Certain Kafka command line tools require direct access to Zookeeper

The following command line tools talk directly to ZooKeeper and therefore are not secured via Kafka:

- kafka-configs
- kafka-reassign-partitions

Workaround:None.

The offsets.topic.replication.factor property must be less than or equal to the number of live brokers

The offsets.topic.replication.factor broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a GROUP_COORDINATOR_NOT_AVAILABLE error until the cluster size meets this replication factor requirement.

Workaround: None.

Requests fail when sending to a nonexistent topic with auto.create.topics.enable set to true

The first few produce requests fail when sending to a nonexistent topic with auto.create.topics.enable set to true.

Workaround: Increase the number of retries in the producer configuration setting retries.

Custom Kerberos principal names cannot be used for kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Workaround: None.

Performance degradation when SSL Is enabled

In some configuration scenarios, significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, Kafka configuration, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with ssl.secure.random.implementation = SHA1PRNG. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Apache JIRA: KAFKA-2561

OPSAPS-43236: Kafka garbage collection logs are written to the process directory

By default Kafka garbage collection logs are written to the agent process directory. Changing the default path for these log files is currently unsupported.

Workaround: None.

CDPD-8546: Repeated ZooKeeper client log trace in Kafka server logs

If the Enable Secure Connection to ZooKeeper property is set to true and the RANGER Service property is configured, both the Kafka Zookeeper client and Ranger Zookeeper client will be configured to connect to Zookeeper via secure channels. However, the Ranger Zookeeper client will try to establish a TLS/SSL connection to an unsecure port (2181). This results in the client repeatedly trying and failing to connect to Zookeeper, which in turn causes the org.apache.zookeeper.Login: TGT refresh thread started. log message as well as other related log messages to repeatedly appear in the Kafka logs.



Note: This issue affects Data Hub clusters only. Data Lake clusters are not affected.

Workaround:

1. In Cloudera Manager, select the Kafka service.
2. Select Configuration and find the Kafka Broker Advanced Configuration Snippet (Safety Valve) for ranger-kafka-audit.xml property.
3. Add the following to the property:

```
name : xasecure.audit.destination.solr.zookeepers
value: ZOOKEEPER_HOST:SECURE_PORT/solr
```

Replace ZOOKEEPER_HOST:SECURE_PORT with the hostname and secure port of the ZooKeeper host that Solr depends on.

4. Enter a Reason for change, and click Save Changes to commit the changes.
5. Restart the service.

Result: The Ranger Zookeeper client connects to the correct secure port.

RELENG-8748: Kafka command line tool alternatives unavailable

The following Kafka command line tool alternatives are currently unavailable:

- kafka-dump-log
- kafka-producer-perf-test
- kafka-streams-application-reset

Workaround: Run the scripts located in /opt/cloudera/parcels/CDH/lib/kafka/bin instead. For example:

```
/opt/cloudera/parcels/CDH/lib/kafka/bin/kafka-producer-perf-test
.sh
```

OPSAPS-57113: The Kafka Broker Advanced Configuration Snippet (Safety Valve) for ssl.properties does not propagate configurations correctly.

If the Kafka Broker Advanced Configuration Snippet (Safety Valve) for ssl.properties property contains configuration that has dollar signs, the configuration is not propagated to Kafka brokers correctly.

Workaround:None.

OPSAPS-59031: Kafka cannot start if configuration is added to the Kafka Broker Advanced Configuration Snippet (Safety Valve) for ssl.properties

The Kafka Broker Advanced Configuration Snippet (Safety Valve) for `ssl.properties` configuration snippet does not correctly override configuration. As a result, Kafka may not start if TLS/SSL related configuration overrides are added to the this configuration snippet.

Workaround: Use the Kafka Broker Advanced Configuration Snippet (Safety Valve) for `kafka.properties` configuration snippet instead to override SSL related properties.

Unsupported Features

The following Kafka features are not supported in Cloudera Data Platform:

- Only Java based clients are supported. Clients developed with C, C++, Python, .NET and other languages are currently not supported.
- Although Kafka Connect is available in the form of a Kafka service role, it is not supported at this time. NiFi is a proven solution for batch and real time data loading that complements Kafka's message broker capability. For more information, see [Cloudera Flow Management](#).
- The Kafka default authorizer is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.

Known Issues in Kerberos

Learn about the known issues in Kerberos, the impact or changes to the functionality, and the workaround.

OPSAPS-60331: If Cloudera Manager is configured to use Active Directory as a Kerberos KDC, and is also configured to use `/etc/cloudera-scm-server/cmf.keytab` as the KDC admin credentials, you may encounter errors when generating Kerberos credentials.

In the Cloudera Manager Admin Console, run the "Administration > Security > Kerberos Credentials > Import KDC Account Manager Credentials" wizard. Remove `/etc/cloudera-scm-server/cmf.keytab` on the Cloudera Manager server host.

Known Issues in Apache Knox

This topic describes known issues and workarounds for using Knox in this release of Cloudera Runtime.

CDPD-3125: Logging out of Atlas does not manage the external authentication

At this time, Atlas does not communicate a log-out event with the external authentication management, Apache Knox. When you log out of Atlas, you can still open the instance of Atlas from the same web browser without re-authentication.

Workaround: To prevent additional access to Atlas, close all browser windows and exit the browser.

OPSAPS-59751: If Cloudera Manager is configured with Apache Knox, then Replication Manager does not work.

None

Technical Service Bulletins

TSB 2022-553: DOM based XSS Vulnerability in Apache Knox

When using Knox Single Sign On (SSO) in the affected releases, a request could be crafted to redirect a user to a malicious page due to improper URL parsing. The request includes a specially crafted request parameter that could be used to redirect the user to a page controlled by an attacker. This request URL would need to be presented to the user outside the normal request flow through a XSS or phishing campaign.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2022-553: DOM based XSS Vulnerability in Apache Knox \("Knox"\)](#)

Known Issues in Apache Kudu

This topic describes known issues and workarounds for using Kudu in this release of Cloudera Runtime.

- Kudu supports only coarse grain authorization
 - Kudu does not yet support integration with Ranger
 - Kudu does not yet support integration with Atlas
- Kudu HMS Sync is disabled and is not yet supported

Known Issues in Apache Oozie

This topic describes known issues and unsupported features for using Oozie in this release of Cloudera Runtime.

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the resume command to inform Oozie to continue the workflow from the point at which it left off.

CDPD-5340: The resourceManager property defined in an Oozie workflow might not work properly if the workflow is submitted through Knox proxy.

An Oozie workflow defined to use the resourceManager property might not work as expected in situations when the workflow is submitted through Knox proxy.

Workaround: Define the jobTracker property with the same value as that of the resourceManager property.

Unsupported Feature

The following Oozie features are currently not supported in Cloudera Data Platform:

- Non-support for Pig action (CDPD-1070)
- Conditional coordinator input logic

Cloudera does not support using Derby database with Oozie. You can use it for testing or debugging purposes, but Cloudera does not recommend using it in production environments. This could cause failures while upgrading from CDH to CDP.

Known Issues in Apache Ranger

This topic describes known issues and workarounds for using Ranger in this release of Cloudera Runtime.

CDPD-3296: Audit files for Ranger plugin components do not appear immediately in S3 after cluster creation

For Ranger plugin components (Atlas, Hive, HBase, etc.), audit data is updated when the applicable audit file is rolled over. The default Ranger audit rollover time is 24 hours, so audit data appears 24 hours after cluster creation.

Workaround:

To see the audit logs in S3 before the default rollover time of 24 hours, use the following steps to override the default value in the Cloudera Manager safety valve for the applicable service.

1. On the Configuration tab in the applicable service, select Advanced under CATEGORY.
2. Click the + icon for the <service_name> Advanced Configuration Snippet (Safety Valve) for ranger-<service_name>-audit.xml property.

3. Enter the following property in the Name box:
xasecure.audit.destination.hdfs.file.rollover.sec.
4. Enter the desired rollover interval (in seconds) in the Value box. For example, if you specify 180, the audit log data is updated every 3 minutes.
5. Click Save Changes and restart the service.

Known Issues in Cloudera Search

This topic describes known issues and unsupported features for using Cloudera Search in this release of Cloudera Runtime.

Known Issues

Splitshard of HDFS index checks local filesystem and fails

When performing a shard split on an index that is stored on HDFS, SplitShardCmd still evaluates free disk space on the local file system of the server where Solr is installed. This may cause the command to fail, perceiving that there is no adequate disk space to perform the shard split.

Workaround: None

Apache Issue: SOLR-14251

Processing UpdateRequest with delegation token throws NullPointerException

When using the Spark Crunch Indexer or another client application which utilizes the SolrJ API to send Solr Update requests with delegation token authentication, the server side processing of the request might fail with a NullPointerException.

Workaround: None.

Apache Issue: SOLR-13921

Solr service with no added collections causes the upgrade process to fail

Upgrade fails while performing the bootstrap collections step of the solr-upgrade.sh script with the error message:

```
Failed to execute command Bootstrap Solr Collections on service Solr
```

if there are no collections present in Solr.

Workaround: If there are no collections added to it, remove the Solr service from your cluster before you start the upgrade.

HBase Lily indexer might fail to write role log files

In certain scenarios the HBase Lily Indexer (Key-Value Store Indexer) fails to write its role log files.

Workaround: None.

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: solr.sentry.enableSqlQuery=true
- Graph: solr.sentry.enableGraphQuery=true
- Stream: solr.sentry.enableStreams=true

Workaround: None.

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search now includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with `CrunchIndexerTool`, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None.

The quickstart.sh file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if `ZooKeeper` and the `NameNode` are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the `NameNode` and `ZooKeeper` are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the `MapReduce` jobs and `Solr` commands that are run by the `quickstart`.

Field value class guessing and Automatic schema field addition are not supported with the MapReduceIndexerTool nor with the HBaseMapReduceIndexerTool

The `MapReduceIndexerTool` and the `HBaseMapReduceIndexerTool` can be used with a `Managed Schema` created via `NRT` indexing of documents or via the `Solr Schema API`. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the `MapReduceIndexerTool` or `HBaseMapReduceIndexerTool`. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the `Solr Schema API` or index sample documents using `NRT` indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect, using the `List Fields API` command.

The Browse and Spell Request Handlers are not enabled in schemaless mode

The `Browse` and `Spell Request Handlers` require certain fields to be present in the schema. Since those fields cannot be guaranteed to exist in a `Schemaless` setup, the `Browse` and `Spell Request Handlers` are not enabled by default.

Workaround: If you require the `Browse` and `Spell Request Handlers`, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. `Blockcache` writing is disabled by default.

Workaround: None.

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given a page explaining that access is denied to them, instead receive a web page that never finishes loading.

Workaround: None.

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None.

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None.

Collection state goes down after Solr SSL

If you enable TLS/SSL on a Solr instance with existing collections, the collections will break and become unavailable. Collections created after enabling TLS/SSL are not affected by this issue.

Workaround:

[Recreate the collection after enabling TLS.](#)

Unsupported Features

The following Search features are currently not supported in Cloudera Data Platform:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation
- Saving search results
- Solr contrib modules (Spark, MapReduce and Lily HBase indexers are not contrib modules but part of the Cloudera Search product itself, therefore they are supported).
- [Logging Slow Queries](#)

Limitations

Default Solr core names cannot be changed

Although it is technically possible to give user-defined Solr core names during core creation, it is to be avoided in the context of Cloudera Search. Cloudera Manager expects core names in the default

"collection_shardX_replicaY" format. Altering core names results in Cloudera Manager being unable to fetch Solr metrics for the given core and this, eventually, may corrupt data collection for co-located core, or even shard and server level charts.

Known Issues in Apache Solr

This topic describes known issues and workarounds for using Solr in this release of Cloudera Runtime.

Technical Service Bulletins

TSB 2021-497: CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler

The Apache Solr ReplicationHandler (normally registered at "/replication" under a Solr core) has a "masterUrl" (also "leaderUrl" alias) parameter. The "masterUrl" parameter is used to designate another ReplicationHandler on another Solr core to replicate index data into the local core. To help prevent the CVE-2021-27905 SSRF vulnerability, Solr should check these parameters against a similar configuration used for the "shards" parameter.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2021-497: CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler](#)

Known Issues in Apache Spark

This topic describes known issues and workarounds for using Spark in this release of Cloudera Runtime.

CDPD-22670 and CDPD-23103: There are two configurations in Spark, "Atlas dependency" and "spark_lineage_enabled", which are conflicted. The issue is when Atlas dependency is turned off but spark_lineage_enabled is turned on.

Run Spark application, Spark will log some error message and cannot continue. That can be restored by correcting the configurations and restarting Spark component with distributing client configurations.

CDPD-217: HBase/Spark connectors are not supported

The Spark HBase Connector (SHC) from HDP and the hbase-spark module from CDH are not supported.

Workaround: Migrate to the Apache HBase Connectors integration for Apache Spark (hbase-connectors/spark) available in CDP. More details on the integration for working with HBase data from Spark in CDP is available in the Cloudera Community article, [HBase and Spark in CDP](#).

CDPD-3038: Launching pyspark displays several HiveConf warning messages

When pyspark starts, several Hive configuration warning messages are displayed, similar to the following:

```
19/08/09 11:48:04 WARN conf.HiveConf: HiveConf of name hive.vectorized.use.checked.expressions does not exist
19/08/09 11:48:04 WARN conf.HiveConf: HiveConf of name hive.tez.cartesian-product.enabled does not exist
```

Workaround: These errors can be safely ignored.

CDPD-2650: Spark cannot write ZSTD and LZ4 compressed Parquet to dynamically partitioned tables

Workaround: Use a different compression algorithm.

CDPD-3293: Cannot create views (CREATE VIEW statement) from Spark

Apache Ranger in CDP disallows Spark users from running CREATE VIEW statements.

Workaround: Create the view using Hive or Impala.

CDPD-3783: Cannot create databases from Spark

Attempting to create a database using Spark results in an error similar to the following:

```
org.apache.spark.sql.AnalysisException:
  org.apache.hadoop.hive.ql.metadata.HiveException: MetaException(message:Permission denied: user [sparkuser] does not have [ALL] privilege on [hdfs://ip-10-1-2-3.cloudera.site:8020/tmp/spark/warehouse/spark_database.db]);
```

Workaround: Create the database using Hive or Impala, or specify the external data warehouse location in the create command. For example:

```
sql("create database spark_database location '/warehouse/tablespace/external/hive/spark_database.db'")
```

Known Issues for Apache Sqoop

This topic describes known issues and workarounds for using Parquet and Avro imports in this release of Cloudera Runtime.

Avro, S3, and HCat do not work together properly

Problem: Importing an Avro file into S3 with HCat fails with Delegation Token not available.

CDPD-3089

Parquet columns inadvertently renamed

Problem: Column names that start with a number are renamed when you use the `--as-parquetfile` option to import data.

Workaround: Prepend column names in Parquet tables with one or more letters or underscore characters.

Apache JIRA: None

Importing Parquet files might cause out-of-memory (OOM) errors

Problem: Importing multiple megabytes per row before initial-page-run check (ColumnWriter) can cause OOM. Also, rows that vary significantly by size so that the next-page-size check is based on small rows, and is set very high, followed by many large rows can also cause OOM.

PARQUET-99

Known Issues in MapReduce and YARN

This topic describes known issues, unsupported features and limitations for using MapReduce and YARN in this release of Cloudera Runtime.

Known Issues

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

CDH-49165: History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

CDH-6808: Routable IP address required by ResourceManager

ResourceManager requires routable host:port addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard `0.0.0.0` address.

Workaround: Set the address, in the form `host:port`, either in the client-side configuration, or on the command line when you submit the job.

OPSAPS-52066: Stacks under Logs Directory for Hadoop daemons are not accessible from Knox Gateway.

Stacks under the Logs directory for Hadoop daemons, such as NameNode, DataNode, ResourceManager, NodeManager, and JobHistoryServer are not accessible from Knox Gateway.

Workaround: Administrators can SSH directly to the Hadoop Daemon machine to collect stacks under the Logs directory.

CDPD-2936: Application logs are not accessible in WebUI2 or Cloudera Manager

Running Containers Logs from NodeManager local directory cannot be accessed either in Cloudera Manager or in WebUI2 due to log aggregation.

Workaround: Use the YARN log CLI to access application logs. For example:

```
yarn logs -applicationId <APPLICATION ID>
```

Apache Issue: [YARN-9725](#)

OPSAPS-50291: Environment variables HADOOP_HOME, PATH, LANG, and TZ are not getting whitelisted

It is possible to whitelist the environment variables `HADOOP_HOME`, `PATH`, `LANG`, and `TZ`, but the container launch environments do not have these variables set up automatically.

Workaround: You can manually add the required environment variables to the whitelist using Cloudera Manager.

1. In Cloudera Manager, select the YARN service.
2. Click the Configuration tab.
3. Search for Containers Environment Variable Whitelist.
4. Add the environment variables (`HADOOP_HOME`, `PATH`, `LANG`, `TZ`) which are required to the list.
5. Click Save Changes.
6. Restart all NodeManagers.
7. Check the YARN aggregated logs to ensure that newly whitelisted environment variables are set up for container launch.

Limitations

Capacity Scheduler

- As Capacity Scheduler is the default scheduler, the Dynamic Resource Pool User Interface is not available by default.
- Capacity Scheduler can be configured only through safety-valves in Cloudera Manager.

COMPX-8687: Missing access check for getAppAttempts

When the Job ACL feature is enabled using Cloudera Manager (YARN Configuration Enable JOB ACL property), the `mapreduce.cluster.acls.enabled` property is not generated to all configuration files, including the `yarn-site.xml` configuration file. As a result the ResourceManager process will use the default value of this property. The default property of `mapreduce.cluster.acls.enabled` is false.

Workaround: Enable the Job ACL feature using an advanced configuration snippet:

1. In Cloudera Manager select the YARN service.
2. Click Configuration.
3. Find the YARN Service MapReduce Advanced Configuration Snippet (Safety Valve) property.
4. Click the plus icon and add the following:
 - Name: mapreduce.cluster.acls.enabled
 - Value: true
5. Click Save Changes.

Unsupported Features

The following YARN features are currently not supported in Cloudera Data Platform:

- GPU support for Docker
- Hadoop Pipes
- Fair Scheduler
- Application Timeline Server (ATS 2 and ATS 1.5)
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Native Services
- Pluggable Scheduler Configuration
- Queue Priority Support
- Reservation REST APIs
- Resource Estimator Service
- Resource Profiles
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- Rolling Log Aggregation
- Docker on YARN (DockerContainerExecutor) on Data Hub clusters
- Moving jobs between queues
- Dynamic Resource Pools

Known Issues in Apache Zeppelin

This topic describes known issues and workarounds for using Zeppelin in this release of Cloudera Runtime.

CDPD-3090: Due to a configuration typo, functionality involving notebook repositories does not work

Due to a missing closing brace, access to the notebook repositories API is blocked by default.

Workaround: From the CDP Management Console, go to Cloudera Manager for the cluster running Zeppelin. On the Zeppelin configuration page (Zeppelin serviceConfiguration), enter shiro urls in the Search field, and then add the missing closing brace to the notebook-repositories URL, as follows:

```
/api/notebook-repositories/** = authc, roles[{{zeppelin_admin_group}}]
```

Click Save Changes, and restart the Zeppelin service.

CDPD-2406: Logout button does not work

Clicking the Logout button in the Zeppelin UI logs you out, but then immediately logs you back in using SSO.

Workaround: Close the browser.

Known Issues in Apache ZooKeeper

This topic describes known issues and workarounds for using Zeppelin in this release of Cloudera Runtime.

Zookeeper-client does not use ZooKeeper TLS/SSL automatically

The command-line tool ‘zookeeper-client’ is installed to all Cloudera Nodes and it can be used to start the default Java command line ZooKeeper client. However even when ZooKeeper TLS/SSL is enabled, the zookeeper-client command connects to localhost:2181, without using TLS/SSL.

Workaround:

Manually configure the 2182 port, when zookeeper-client connects to a ZooKeeper cluster. The following is an example of connecting to a specific three-node ZooKeeper cluster using TLS/SSL:

```
CLIENT_JVMFLAGS="-Dzookeeper.clientCnxnSocket=org.apache.zoo
keeper.ClientCnxnSocketNetty -Dzookeeper.ssl.keyStore.locati
on=<PATH TO YOUR CONFIGURED KEYSTORE> -Dzookeeper.ssl.keyStor
e.password=<THE PASSWORD YOU CONFIGURED FOR THE KEYSTORE> -
Dzookeeper.ssl.trustStore.location=<PATH TO YOUR CONFIGURED
TRUSTSTORE> -Dzookeeper.ssl.trustStore.password=<THE PASSWORD
YOU CONFIGURED FOR THE TRUSTSTORE> -Dzookeeper.client.secu
re=true" zookeeper-client -server <YOUR.ZOOKEEPER.SERVER-1>:218
2,<YOUR.ZOOKEEPER.SERVER-2>:2182,<YOUR.ZOOKEEPER.SERVER-3>:2182
```