

## Cloudera DataFlow for Data Hub Release Notes

Date published: 2022-02-24

Date modified: 2022-02-24



# Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

# Contents

<b>What's New in Cloudera DataFlow for Data Hub 7.2.14.....</b>	<b>4</b>
What's New in Flow Management.....	4
What's New in Streams Messaging.....	4
Cruise Control Rebase Summary.....	10
What's New in Streaming Analytics.....	12
 <b>CVE-2021-45105 &amp; CVE-2021-44832 remediation for CDF for Data Hub.....</b>	 <b>13</b>
 <b>Component Support in Cloudera DataFlow for Data Hub 7.2.14.....</b>	 <b>14</b>
 <b>Supported NiFi Extensions.....</b>	 <b>14</b>
Supported NiFi Processors.....	14
Supported NiFi Controller Services.....	17
Supported NiFi Reporting Tasks.....	18
Components Supported by Partners.....	18
 <b>Unsupported Features in Cloudera DataFlow for Data Hub 7.2.14.....</b>	 <b>19</b>
Unsupported Flow Management features.....	19
Unsupported Streams Messaging features.....	19
Unsupported Streaming Analytics features.....	20
 <b>Known Issues In Cloudera DataFlow for Data Hub 7.2.14.....</b>	 <b>21</b>
Known Issues in Flow Management.....	21
Known Issues in Streams Messaging.....	22
Known Issues in Streaming Analytics.....	28
 <b>Fixed Issues in Cloudera DataFlow for Data Hub 7.2.14.....</b>	 <b>30</b>
Fixed Issues in Flow Management.....	30
Fixed Issues in Streams Messaging.....	36
Fixed Issues in Streaming Analytics.....	38

## What's New in Cloudera DataFlow for Data Hub 7.2.14

Cloudera DataFlow for Data Hub 7.2.14 includes components for Flow Management, Streaming Analytics, and Streams Messaging. Learn about the new features and improvements in each of these components.

### What's New in Flow Management

Learn about the new Flow Management features in Cloudera DataFlow for Data Hub 7.2.14.

This release of Flow Management is based on the latest version of Apache NiFi 1.15 and includes significant improvements. Here are some important new features:

- Improved NiFi UI responsiveness  
The experience when using NiFi (UI) with flows containing thousands or tens of thousands of components is improved (NIFI-9309).
- Parameter Contexts inheritance  
Provides the ability to leverage the Parameter Contexts in a more flexible way in multi-tenant environments (NIFI-8490).
- Improved Hadoop-related processors class loading  
Improves how class loading isolation works around the UserGroupInformation class for Hadoop-related components. This improves startup time for NiFi when using thousands of these components (NIFI-9382).
- ExecuteStateless processor  
The option to execute flows in NiFi while using the Stateless implementation. Improves performance for specific flows as well as providing exactly-once operations in some cases (NIFI-9239).
- Components validation  
Provides a way to validate and get insights about a component's configuration in NiFi's (UI) configuration view (NIFI-9009).
- New components  
New processors for ElasticSearch, new Scripted processors to better support custom business logic while processing data, new processors to use PGP for signing/verifying content, a new email record sink for monitoring purposes to be used with reporting tasks.
- Apache Pulsar integration in partnership with StreamNative  
New processors and controller service to interact with Apache Pulsar. These new components are built, supported and maintained by StreamNative.

### What's New in Streams Messaging

Learn about the new Streams Messaging features in Cloudera DataFlow for Data Hub 7.2.14.

#### Kafka

##### Rebase on Kafka 2.8.0

Kafka shipped with this version of Cloudera Runtime is based on Apache Kafka 2.8.0. For more information, see the following upstream resources:

Apache Kafka Notable Changes:

- [2.6.0](#)
- [2.7.0](#)
- [2.8.0](#)

Apache Kafka Release Notes:

- [2.6.0](#)
- [2.7.0](#)
- [2.8.0](#)

### **Kafka broker rolling restart checks**

Cloudera Manager can now be configured to perform different types of checks on the Kafka brokers during a rolling restart. Using these checks can ensure that the brokers remain healthy during and after a rolling restart. As a result of this change, Kafka rolling restarts may take longer than in previous versions. This is true even if you disable the rolling restart checks. For more information, see [Rolling restart checks](#).

### **Http Metrics Report Exclude Filter introduced for Kafka**

A new property, Http Metrics Report Exclude Filter (`kafka.http.metrics.reporter.exclude.filter`), is introduced for the Kafka service. This property can be used to specify a regular expression that is used to filter metrics. Any metric matching the specified regular expression is not reported by Cloudera Manager. As a result, these metrics are also not displayed in SMM. Use JMX metric names when configuring this property.

### **Bootstrap servers are automatically configured for Kafka Connect**

The Bootstrap Servers property of the Kafka Connect role is now automatically configured to include the bootstrap servers of its co-located Kafka brokers. This is only done if the property is left empty (default). You can provide custom value for this property if you want to override the default host:port pairs that Kafka Connect uses when it establishes a connection with the Kafka brokers.

### **Kafka Connect Ranger Authorizer**

A Ranger plugin is introduced for Kafka Connect that implements the Authorizer interface. A new service type is now also introduced in Ranger called `kafka-connect`. By default it includes the `cm_kafka_connect` resource-based service which includes policies that provide default access. The default resource-based service that is created for Kafka Connect can be configured using the 'Ranger service' name for the Kafka Connect service (`ranger_plugin_kafka_connect_service_name`) Kafka service property.

### **Kafka Connect in DataHub [Technical Preview]**

Kafka Connect can now be provisioned in CDP Public Cloud with Data Hub. The default Streams Messaging cluster definitions are updated to include Kafka Connect. For more information, see [Streams Messaging cluster layout](#), [Creating your first Streams Messaging cluster](#), and [Scaling Kafka Connect](#).

### **Stateless NiFi Source and Sink [Technical Preview]**

The Stateless NiFi Source and Sink connectors enable you to run NiFi dataflows within Kafka Connect. Using these connectors can grant you access to a number of NiFi features without having the need to deploy or maintain NiFi on your cluster. For more information on the connectors, best practices on building dataflows to use with these connectors, as well as information on how to deploy the connectors, see [Stateless NiFi Source and Sink](#).

### **New Cloudera developed Kafka Connect connectors [Technical Preview]**

In addition to the introduction of the Stateless NiFi Source and Sink, 12 new Cloudera developed connectors are available for use with Kafka Connect. These are powered by the Stateless NiFi engine and run Cloudera developed dataflows. They provide an out-of-the box solution for some of the most common use cases for moving data in or out of Kafka. For more information, see [Connectors](#) in the Kafka Connect documentation.

### **Kafka multiple Availability Zone support [Technical Preview]**

Kafka can now be deployed in multiple Availability Zones in CDP Public Cloud. When using the multi Availability Zone feature, CDP ensures that Kafka replicates partitions across brokers in different availability zones. For more information, see [Deploying CDP in multiple AWS availability zones](#).

**The default replication factor for Kafka Connect internal topics is increased to 3**

The default replication factor for Kafka Connect internal topics was set to 1. The replication factor is increased to 3. The following properties are affected:

- Offset Storage Topic Replication Factor(offset.storage.replication.factor)
- Configuration Storage Topic Replication Factor(config.storage.replication.factor))
- Status Storage Topic Replication Factor (status.storage.replication.factor)

**Schema Registry****Support added for JSON schemas in Schema Registry**

The JSON type schema format is now supported.

Earlier, only Avro type schema format was supported out of the box.

**Cloudera Manager supports rolling restarts of HA enabled Schema Registry**

Schema Registry service can now be rolling restarted using Cloudera Manager.

**Added Import tool for Schema Registry schemas**

Schemas stored in Schema Registry can be exported to a JSON file. The exported JSON file can then be imported into another Schema Registry database. During an import, SchemaMetadata, SchemaBranch, and SchemaVersion objects are put into the database. These objects retain their ID as well as a number of other properties that are available in the JSON file used for import. This way, serializing and deserializing protocols can continue to function without any change and Schema Registry clients can seamlessly switch between different Schema Registry instances. Both import and export operations are done using the Schema Registry API.

**Streams Messaging Manager****Reactive Lineage fetching from Kafka producer cache**

You can now visualize the lineage between producers and consumers in SMM. Lineage information helps you to understand how the message is moving from a producer to a consumer group and which topics or partitions are part of that flow. Lineage between clients and topics or partitions are now shown using the new lineage endpoints. For more information, see [Monitoring lineage information](#).

**New endpoint added to fetch lineage for a topic**

The `/api/v1/admin/lineage/partitions/{topic}` endpoint used to fetch which producers have produced into the queried topic, and which consumerGroup's members have consumed from it. Now when you click on a topic to fetch the lineage on the UI, this endpoint is used.

**New endpoint added to fetch lineage information for a consumerGroup**

The `/api/v1/admin/lineage/consumerGroups/{consumerGroupId}` endpoint used to fetch which topics the members of that consumerGroup have consumed from, and also what producers have produced into those topics. Now when you click on a group on the UI to fetch the lineage, this endpoint is used.

**New endpoint added to fetch lineage information for a topicPartition**

The `/api/v1/admin/lineage/partitions/{topic}/{partition}` endpoint is used to fetch which producers have produced into that queried topicPartition and which consumerGroup members have consumed from that topicPartition. Now when you click on a topicPartition to fetch the lineage on the UI, this endpoint is used.

**New endpoint added to fetch lineage information for a producer**

The `/api/v1/admin/lineage/lineage/producers/{producerId}` endpoint is used to fetch which topics the queried producer has produced into, and which consumerGroups members have consumed from those topics. Now when you click on a producer to get the lineage on the UI, this endpoint is used.

**On selecting the partition on Overview page, the new lineage endpoint should be called**

Lineage between clients and topics or partitions are now shown using the new lineage endpoints. Remember that when checking the lineage (connected clients) for a TopicPartition, only the recently connected clients will be shown.

### **Support added for multiple replication targets in SMM**

SMM now supports SRM replication flows targeting remote clusters making use of the new v2 SRM APIs.

Remote Replication flows available under the `/api/v2/admin/replication-stats` APIs. The UI is now configured to make use of these new APIs.

### **Introduced multi-target replication monitoring support in alerts**

- SMM now adopted the new V2 SRM Service endpoints upon which alerting is based on.
- In SMM when configuring alerts for replications in the UI now source and target clusters can be defined, as opposed to the previous configuring panel, where only the source cluster could be defined (since the target cluster was fixed to be the colocated Kafka cluster).
- Old alerts will still function, however editing them can only be done using the new format, where source and target clusters have to be defined.
- **IMPORTANT:** For alerts involving remote SRM cluster queries set the execution interval to at the very least a minute (preferably more).

### **SMM authenticates to SRM Service**

SMM now automatically configures Basic Authentication when connecting to SRM and the service dependency based auto-configuration is in use.

For manual SRM connectivity configurations, Basic Auth configurations were added (Streams Replication Manager Basic Authentication, Streams Replication Manager Basic Authentication Username, Streams Replication Manager Basic Authentication Password).

### **SMM Cache-Control is part of default SMM REST Server API's responses' headers**

The new SMM configuration named `cache.control.http.response.header.value` allows to configure the Cache-Control header's value for certain endpoints. Configure it in the following key-value like fashion:

- The key is the path prefix to the endpoints where the Cache-Control header should be added.
- The value is the value of Cache-Control header.

In order to turn off functionalities provided by the Cache-Control header just delete the entries, or set the value to no-store.

### **Added helper tooltips to SMM UI**

SMM now provides more informative tooltips (hover over the table headers and labels) for most of its elements in the web UI.

### **Removed Consumer Rate graphs**

The lag rate graph is removed from the UI.

The lag rate values are removed from the `/api/v1/admin/metrics/aggregated/groups/{groupName}` and `/api/v1/admin/metrics/aggregated/groups` endpoints.

### **SMM is automatically integrated with co-located Kafka Connect**

To monitor and manage Kafka Connect in SMM, a number of SMM service properties must be configured. These are the following:

- Kafka Connect Host
- Kafka Connect Port
- Kafka Connect Protocol

From now on, these properties are automatically configured if Kafka Connect Host is left empty (default). This means that the SMM service automatically configures itself to connect to its co-

located Kafka Connect instance. You can provide custom values for all properties if you want to override the defaults.

## Streams Replication Manager

### SRM Driver monitoring using Cloudera Manager

Cloudera Manager's ability to monitor the SRM Driver, its replications, and the overall health of SRM is improved. Most notably, the health status of SRM is based on the health of the network and the availability of replication sources and targets. As a result of this improvement, two new metrics, a new health test, and several new configuration properties are introduced for the SRM Driver in Cloudera Manager.

#### New metrics and health test

The new metrics are as follows:

- SRM Driver Distributed Herder Status (`streams_replication_manager_distributed_herder_status`)
- Aggregated Status Code of SRM Driver Replication Flows (`streams_replication_manager_aggregated_herder_status`)

The distributed metric describes the status of individual replications. The aggregate metric provides the aggregate status of all replications.

The new health test is called `DISTRIBUTED_HERDER_STATUS`. This health test is based on the aggregate metric and provides information on the overall status of SRM and its replications.

#### New properties

The new monitoring related properties are as follows:

- Path for driver plugins (`plugin.path`)
- Enable HTTP(S) Metrics Reporter (`mm.metrics.servlet.enable`)
- SSL Encryption for the Metrics Reporter (`metrics.jetty.server.ssl.enabled`)
- HTTP Metrics Reporter Port (`metrics.jetty.server.port`)
- HTTPS Metrics Reporter Port (`metrics.jetty.server.secureport`)
- Enable Basic Authentication for Metrics Reporter (`metrics.jetty.server.authentication.enabled`)
- Metrics Reporter User Name (`metrics.jetty.server.auth.username`)
- Metrics Reporter Password (`metrics.jetty.server.auth.password`)

For more information, see [Cloudera Manager Configuration Properties Reference](#).

### SRM Driver now automatically retries setting up replications for unavailable target Kafka clusters

Previously, if any of the Kafka clusters that were targeted by the SRM Driver were unavailable at startup, the SRM Driver stopped. As a result of an improvement, the SRM Driver now instead sets up replications for all target Kafka clusters that are available and continuously retries to set up replication for unavailable clusters. Retry behaviour is configurable in Cloudera Manager. The new properties related to retry behaviour are as follows:

- Retry Count for SRM Driver (`mm.replication.restart.count`)
- Retry Delay for SRM Driver (`mm.replication.restart.delay.ms`)

For more information see, [Cloudera Manager Configuration Properties Reference](#) or [Configuring SRM Driver retry behaviour](#).

### Disabled replications can now be fully deactivated by configuring heartbeat emission

As a result of the rebase to Kafka 2.8 (KAFKA-10710), an improvement is introduced in connection with heartbeat emission. From now on, you can fine tune your deployment and fully deactivate any unnecessary replications that are set up by default by configuring heartbeat emission. This can help with minimizing any performance overhead caused by unnecessary replications.

To support this change, an improvement was made for the SRM service in Cloudera Manager. A dedicated configuration property, `Enable Heartbeats`, is introduced. You can use this property to configure `emit.heartbeats.enabled` on a global level directly in Cloudera Manager. Replication



level overrides are still supported. This can be done by adding `emit.heartbeats.enabled` with a valid replication prefix to Streams Replication Manager's Replication Configs. For more information on configuring heartbeat emission, see [Configuring SRM Driver heartbeat emission](#).

### IdentityReplicationPolicy now available



**Warning:** The IdentityReplicationPolicy does not detect cycles. As a result, using this replication policy is only viable in deployments where the replication setup is acyclic. If your replication setup is not acyclic, using this replication policy might result in records being replicated in an infinite loop between clusters. Additionally, monitoring replications with the SRM Service is not possible when this policy is in use.

The version of Apache Kafka shipped with this release of Cloudera Runtime includes KAFKA-9726. As a result, the IdentityReplicationPolicy is available for use with Streams Replication Manager. This replication policy does not rename remote (replicated) topics. Streams Replication Manager can be configured to use this replication policy by adding the following entry to Streams Replication Manager's Replication Configs:

```
replication.policy.class=org.apache.kafka.connect.mirror.IdentityReplicationPolicy
```

For more information, see [KAFKA-9726](#).

### SRM configuration properties can be configured globally for Connect workers and Connect connectors

The SRM Driver now accepts configuration properties prefixed with the `workers.` and `connectors.` prefixes. Configuration properties added to Streams Replication Manager's Replication Configs that use these prefixes are applied globally to all Connect workers or Connect connectors that the SRM Driver creates. For more information regarding the prefixes, see [Understanding SRM properties, their configuration and hierarchy](#). For more information on Connect workers and connectors, see [Streams Replication Manager Architecture](#).

### SRM Service Basic Authentication support

The SRM Service can now be secured using Basic Authentication. Once Basic Authentication is set up and enabled, the REST API of the SRM Service becomes secured. Any clients or services that connect to the REST API will be required to present valid credentials for access. Configuration is done in Cloudera Manager using SRM configuration properties and external accounts. For more information, see [Configuring Basic Authentication for the SRM Service](#).

### SRM automatically creates a Basic Authentication credential for co-located services

SRM automatically creates a Basic Authentication credential for co-located services (users can change the credentials using SRM Service Co-Located Service Username and SRM Service Co-Located Service User Password). When Basic Authentication is enabled, this user is automatically accepted by the SRM Service. For more information, see [Configuring Basic Authentication for the SRM Service](#).

### SRM Service Remote Querying no longer in technical preview

SRM Service Remote Querying was introduced in a previous release of Cloudera Runtime as a technical preview feature. Starting with this release, Remote Querying is ready for use in production environments. This is the result of Basic Authentication being introduced for the SRM Service and SMM supporting multi-target alerting.

For more information on Remote Querying, see [Remote Querying](#) and [Configuring Remote Querying](#). For more information on how to set up Basic Authentication for Remote Querying, see [Configuring Basic Authentication for Remote Querying](#).

### The SRM Driver can now write the origin offset into the record header

SRM now supports a diagnostic feature in which the source offset of the replicated records are written into the headers. The feature can be turned on by setting `copy.source.offset.in.header.enabled` to `true`. When enabled, the source offset is written into a header named `mm2-source-offset` in

binary format. The schema of the header payload is available in the connect:mirror-client package, the class name is `org.apache.kafka.connect.mirror.SourceOffsets`. This feature is only recommended for diagnostic purposes, as the header change increases the size of the replica topic.

## Cruise Control

### Cruise Control 2.5.66 Rebase

Cruise Control in Cloudera Runtime is rebased from 2.0.100 to the 2.5.66 version. The main feature changes include ZooKeeper TLS/SSL support and the Cruise Control Metric Reporter support.

### ZooKeeper TLS/SSL support for Cruise Control

When TLS is enabled on the cluster, Cruise Control automatically uses the Zookeeper for secure communication.

### Cruise Control Metric Reporter support

Beside the Cloudera Manager Metrics Reporter, the Kafka based Cruise Control Metrics Reporter can also be used. The configuration needs to be set manually, and further adjustments are needed when changing the default Metrics Reporter.

### Replacing RackAwareGoal to RackAwareDistributionGoal

The default RackAwareGoal is too strict to allow multiple replicas of the same partition to be placed into a single rack. This means that if there is an outage in one of the availability zones, Cruise Control cannot execute the reassignment. Using the RackAwareDistributionGoal improves this functionality of Cruise Control, as it allows multiple replicas of a partition to be placed into a single rack. This can be achieved as long as the replicas of each partition can be evenly distributed across the racks.

## Cruise Control Rebase Summary

In CDP Public Cloud 7.2.14, Cruise Control is rebased from 2.0.100 to the 2.5.66 version. Other than the added new feature, several issues are fixed and several features are enhanced to have a better performance when using Cruise Control.

**Table 1: Fixed Issues**

<a href="#">PR-1184</a> Fix the bug in replica movement strategy chaining	<a href="#">PR-1291</a> Fix GOAL_VIOLATION detector getting stuck execution in GENERATING_PROPOSALS_FOR_EXECUTION state
<a href="#">PR-1209</a> Fix NPE if task execution takes longer than executionProgressCheckIntervalMs()	<a href="#">PR-1381</a> Fix a bug that might cause invalid throttle replica list to be used
<a href="#">PR-1231</a> Fix reported balancedness when there are offline brokers	<a href="#">PR-1476</a> Fix miscalculated recommendation for the number of racks to drop for clusters over-provisioned wrt rack count
<a href="#">PR-1232</a> Fix returns for completed_with_error tasks on user_tasks endpoint when json=false	<a href="#">PR-1597</a> Fix incorrect values generated for number of replicas by topic
<a href="#">PR-1238</a> Fix inconsistent/bad response in monitor substate	<a href="#">PR-1616</a> Fix throttler quota removal for in-progress tasks
<a href="#">PR-1279</a> Fix missed broker failure detection/self-healing upon bootstrap	<a href="#">PR-1676</a> Fix EnvConfigProvider to work well if there is no pre configured env vars

**Table 2: Version Update**

<a href="#">PR-1233</a> Upgrade to Kafka 2.5.0 in development branch migrate_to_kafka_2_5
<a href="#">PR-1311</a> Add support for Kafka 2.6 brokers
<a href="#">PR-1471</a> Add support for Kafka version 2.7
<a href="#">PR-1612</a> Upgrade to Kafka 2.8 libraries
<a href="#">PR-1614</a> Updated to Scala 2.13

**Table 3: Feature Support**

<a href="#">PR-1159</a> Support SPNEGO and trusted proxy authentication	<a href="#">PR-1569</a> Add support to switch from ZK to Kafka Admin Client for topic config provider class
<a href="#">PR-1245</a> , <a href="#">PR-1320</a> Bump up ZK session and connection timeout	<a href="#">PR-1583</a> Ability to configure TLS protocols and ciphers via configuration
<a href="#">PR-1510</a> Add support for Alerta.io notifications	<a href="#">PR-1661</a> Support connecting to ZooKeeper with TLS
<a href="#">PR-1525</a> Add support for (At/Under)MinISR-based throttling/cancellation	<a href="#">PR-1703</a> Add ZK TLS support with properties file and modify broker failure detection

**Table 4: Goal Improvements**

<a href="#">PR-1203</a> Add missing sanity check for goals	<a href="#">PR-1400</a> Add gap-based balance limits for TopicReplicaDistributionGoal
<a href="#">PR-1267</a> LeaderReplicaDistributionGoal should honor excludedTopics during leadership movement	<a href="#">PR-1420</a> Relax low resource utilization upper limit for resource distribution goals
<a href="#">PR-1306</a> Update min valid windows required to start self-healing with goals using resource history	<a href="#">PR-1429</a> Add a new hard goal that ensures that each alive broker has a leader replica from a configured pool of topics
<a href="#">PR-1324</a> Add support for goal-based operations via maintenance event	<a href="#">PR-1500</a> Drop enforcement that anomaly.detection.goals must be a subset of self.healing.goals (if non-empty)
<a href="#">PR-1345</a> Add a new hard goal that evenly distributes replicas over racks	<a href="#">PR-1514</a> Prevent ResourceDistributionGoal from generating provision recommendations with a negative number of brokers to remove
<a href="#">PR-1383</a> Ensure that topology distribution goals compute balance constraints properly	<a href="#">PR-1564</a> Add timers to track goal violation detection and fix generation for self-healing
<a href="#">PR-1385</a> Add support to switch from ZK to Kafka Admin Client for topic config provider class	

**Table 5: Functionality Enhancements**

<a href="#">PR-868</a> Provide capacity stats for a broker	<a href="#">PR-1302</a> Support stop ongoing executions with rollback	<a href="#">PR-1448</a> Reset the provision status to ensure freshness for consecutive optimizations
<a href="#">PR-1177</a> Make TopicReplicationFactorAnomalyFinder ignore topic with large minISR	<a href="#">PR-1313</a> Make min execution progress check interval and slow task alerting backoff configurable	<a href="#">PR-1456</a> Update slow broker detection sensitivity and reporting details
<a href="#">PR-1180</a> Update min valid windows required to start self-healing with goals using resource history	<a href="#">PR-1316</a> Added option to configure metrics topic minISR	<a href="#">PR-1460</a> Add a config for admin client request timeout
<a href="#">PR-1186</a> Further filter detected slow broker against pre-defined flush time threshold	<a href="#">PR-1332</a> Support handling planned maintenance events submitted via a topic	<a href="#">PR-1463</a> Add sensors to report the number of slow brokers
<a href="#">PR-1190</a> [ccclient-1.1.0] Add force_stop parameter	<a href="#">PR-1334</a> [ccclient-1.1.1] Extend support for anomaly detectors	<a href="#">PR-1469</a> Report ongoing replica reassignments started by an external agent
<a href="#">PR-1196</a> Adopt admin client-based replica reassignment API for Kafka 2.4+	<a href="#">PR-1341</a> Adopt a shared AdminClient across selected CC components	<a href="#">PR-1470</a> Provide recommendations on the estimated resource requirements
<a href="#">PR-1198</a> Support environment variable resolution in configs	<a href="#">PR-1349</a> Make MetricSampler more extensible	<a href="#">PR-1484</a> Add a sensor to indicate the metadata factor of the managed Kafka cluster
<a href="#">PR-1199</a> Add rack information to load response	<a href="#">PR-1357</a> Add check to validate that time range start time is smaller than end time	<a href="#">PR-1485</a> Add a sensor to indicate if the cluster has partitions with RF > the number of eligible racks
<a href="#">PR-1212</a> Make CPU capacity threshold stricter	<a href="#">PR-1358</a> Check whether a cluster is using JBOD when populate_disk_info is true	<a href="#">PR-1496</a> Stop execution before shutting down Cruise Control
<a href="#">PR-1214</a> Update the Balancedness Score under Unhealthy Cluster State	<a href="#">PR-1360</a> Handle Wrapped AdminClient Timeouts as Timeouts	<a href="#">PR-1505</a> Make CC inter-broker replica reassignments resilient against partitions with ISR set > replica set

<a href="#">PR-1220</a> Enable configurable backoff on metrics topic creation	<a href="#">PR-1362</a> Prevent MaintenanceEventTopicReader from prematurely closing the adminClient	<a href="#">PR-1509</a> Add a sensor to identify if the cluster has partitions with ISR > replicas
<a href="#">PR-1223</a> Ensure that requests to update replication factor cannot cause a deadlock	<a href="#">PR-1364</a> Let implementations of OptimizationOptionsGenerator be configured with an AdminClient	<a href="#">PR-1533</a> Avoid NPE due to misused rebalance_disk parameter
<a href="#">PR-1225</a> Allow customization of CruiseControlMetricsReporterSampler	<a href="#">PR-1368</a> Automate leadership concurrency adjustment based on broker metrics	<a href="#">PR-1538</a> Enable partition metric collection to a configured topic during ongoing execution
<a href="#">PR-1234</a> Ensure consistent rackID during topic_configuration operations to avoid NPE	<a href="#">PR-1369</a> Enable Cruise Control to collect metrics from low traffic clusters by default	<a href="#">PR-1559</a> Add a ReplicaMovementStrategy that prioritizes (At/Under)MinISR partitions with offline replicas
<a href="#">PR-1241</a> Add support to retrieve capacity only via load endpoint	<a href="#">PR-1391</a> Calculate balance lower bound for resource distribution lower bound with low utilization threshold	<a href="#">PR-1589</a> Add parameters to relevant endpoints to control the speed of proposal generation
<a href="#">PR-1246</a> Enable Kafka port retrieval from listeners config	<a href="#">PR-1401</a> Honor webserver.api.urlprefix config	<a href="#">PR-1593</a> Set the default proposal generation speed to fast mode
<a href="#">PR-1255</a> Enable broker metric collection during ongoing executions	<a href="#">PR-1407</a> Provide idempotency support to handle duplicate maintenance events	<a href="#">PR-1599</a> Add a config to enable/disable provisioner
<a href="#">PR-1256</a> Prevent Cruise Control from mistakenly believe that there is an ongoing execution	<a href="#">PR-1409</a> [ccclient-1.1.2] Add topic parameter in cruise-control-client to query the kafka_cluster_state endpoint	<a href="#">PR-1604</a> Move broker failure detector away from using zNode-based failed broker persistence
<a href="#">PR-1282</a> Let executor substate show information on process before starting an execution	<a href="#">PR-1410</a> Add a sensor to emit topic count in cluster	<a href="#">PR-1608</a> Disable bootstrap endpoint in non-developer_mode
<a href="#">PR-1287</a> Prevent concurrent execution request from corrupting ongoing execution state	<a href="#">PR-1412</a> Enable detecting and fixing maintenance events during ongoing executions	<a href="#">PR-1622</a> Enable users to monitor the min.insync.replicas of all topics
<a href="#">PR-1289</a> Automate replica reassignment concurrency adjustment based on broker metrics	<a href="#">PR-1418</a> Add missing population of anomaly details for maintenance events	<a href="#">PR-1635</a> Add config to skip rack-awareness check while self-healing RF anomalies
<a href="#">PR-1294</a> Add sensors to emit concurrency caps for partition and leadership reassignments	<a href="#">PR-1419</a> Handle non-existent topic while setting/removing throttled replicas for a topic	<a href="#">PR-1636</a> Add capability to stop executions not started by Cruise Control
<a href="#">PR-1297</a> Make timeout for listing partition reassignments configurable along with a retry logic	<a href="#">PR-1433</a> Handle missing listeners config in CruiseControlMetricsReporter	<a href="#">PR-1637</a> Detect RF violations for topics having targetReplicationFactor with topicReplicationFactorMargin violation
<a href="#">PR-1646</a> Setup ProvisionerState for rightsizing clusters	<a href="#">PR-1681</a> Handle metrics reporter exceptions while getting CPU metric	

## What's New in Streaming Analytics

Learn about the new Streaming Analytics features in Cloudera DataFlow for Data Hub 7.2.14.

The following new features are introduced in Streaming Analytics CDF for Data Hub 7.2.14:

### Apache Flink upgrade

Apache Flink 1.14 is supported in Streaming Analytics 7.2.14 cluster definitions.

For more information on what is included in the Apache Flink 1.14 version, see the [Apache Flink 1.14 Release Post](#) and the [Apache Flink 1.14 Release Notes](#).

### Db2 CDC connector support

Db2 CDC connector is added for the set of supported connectors. This enables you to use data from the IBM Db2 databases with the Change Data Capture connector in SSB.

For more information, see the [CDC connectors documentation](#).

### Custom connectors and data formats

You are able to add new connectors and data formats, and also modify the default and predefined connectors and data formats using Streaming SQL Console to further customize your Streaming SQL jobs.

For more information, see the *Supported connectors* and *Managing connectors* in the [Connectors](#) sections.

### SQL job migration tool

With the SQL job migration tool you can migrate your SQL jobs with configurations from one cluster to another using REST API or Command Line Interface (CLI).

For more information, see the [Using the SQL job migration tool](#) section.

### REST API support

You can use the SQL Stream Builder (SSB) REST API to manage and monitor your SQL jobs, sessions and queries from the CLI using POST, GET and DELETE endpoints. There is also a direct link for the REST API Explorer from Cloudera Manager and from Streaming SQL Console as well.

For the list of newly added operations and more information about the SSB REST API, see the [Using the SSB REST API](#) section.

### Flink Dashboard access

Flink Dashboard is added to the Streaming SQL Console main menu for easier job monitoring access. The SSB Team authorization feature is also extended to Flink Dashboard which means that only those jobs can be monitored in Flink Dashboard that belong to the team a user is assigned to.

### Auto discovery of services

Using the auto discovery of services, you can automatically import data providers and catalogs to SQL Stream Builder from your clusters in your environment.

For more information, see the [Using auto discovery of services](#) section.



**Important:** The Auto discovery of services is provided as a technical preview at this time. The tool is still under development and not recommended for a production environment.

## CVE-2021-45105 & CVE-2021-44832 remediation for CDF for Data Hub

Learn more about the CVE-2021-45105 and CVE-2021-44832 remediation for the Flow Management, Streams Messaging and Streaming Analytics cluster templates in CDF for Data Hub.

On February 1, 2022, Cloudera released a hotfix to Public Cloud Runtime version 7.2.12. It addresses the CVE and other vulnerability concerns as listed below:

- [CVE-2021-45105](#) which affects Apache Log4j2 versions from 2.0-beta9 to 2.16.0, excluding 2.12.3
- [CVE-2021-44832](#) which affects Apache Log4j2 versions from 2.0-alpha7 to 2.17.0, excluding 2.3.2 and 2.12.4

The following table summarizes which template is impacted by the vulnerabilities:

Template	Impacted versions
Flow Management	All versions
Streams Messaging	Not impacted
Streaming Analytics	All versions from 7.2.10

As the CDF for Data Hub cluster templates are running in the CDP Public Cloud environment powered by Runtime, Cloudera encourages users to upgrade their CDP services running Runtime versions from 7.2.7 so that they include

the latest hotfixes. You can update your existing Data Lake and Data Hubs by doing a maintenance upgrade. For more information, see the [Data Lake upgrade](#) and [Data Hub upgrade](#) documentation.



**Note:** Maintenance upgrades are not supported for RAZ-enabled environments.

If you are running a version of Runtime lower than 7.2.7, contact Cloudera Support for details on how to upgrade Runtime.

For more information about the impacts of CVE-2021-45105, see the [TSB 2021-547: Critical vulnerability in log4j2 CVE-2021-45105 Knowledge Base article](#).

## Component Support in Cloudera DataFlow for Data Hub 7.2.14

Cloudera DataFlow for Data Hub 7.2.14 includes the following components.

Flow Management clusters

- Apache NiFi 1.15.2.2.2.4.0
- Apache NiFi Registry 1.15.2.2.2.4.0



**Note:** Apache NiFi and Apache NiFi Registry version are unified in the 1.15.x release.

Streams Messaging clusters

- Apache Kafka 2.8.1
- Schema Registry 0.10.0
- Streams Messaging Manager 2.2.0
- Streams Replication Manager 1.1.0
- Cruise Control 2.5.66

Streaming Analytics clusters

- Apache Flink 1.14

## Supported NiFi Extensions

Apache NiFi 1.15.2 ships with a set of Processors, Controller Services, and Reporting Tasks, most of which are supported by Cloudera Support. Review the supported extensions and avoid using any unsupported extensions in your production environments.

## Supported NiFi Processors

This release ships with Apache NiFi 1.15.2 and includes a set of Processors, most of which are supported by Cloudera Support. You should be familiar with the available supported Processors, and avoid using any unsupported Processors in production environments.

Additional Processors are developed and tested by the Cloudera community but are not officially supported by Cloudera. Processors are excluded for a variety of reasons, including insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, and feature deviation from Cloudera best practices.

AttributesToCSV

GetFTP

PutElasticsearch 1

AttributesToJSON	GetHBase	PutElasticsearchHttp 1
Base64EncodeContent	GetHDFS	PutElasticsearchHttpRecord
CalculateRecordStats	GetHDFSFileInfo	PutElasticsearchRecord
CaptureChangeMySQL	GetHDFSSequenceFile	PutEmail 1
CompressContent12	GetHTMLElement	PutFile
ConnectWebSocket	GetHTTP	PutFTP
ConsumeAMQP	GetIgniteCache	PutGCSObject
ConsumeAzureEventHub	GetJMSQueue	PutGridFS
ConsumeEWS	GetJMSTopic	PutHBaseCell 1
ConsumeGCPubSub	GetKafka	PutHBaseJSON
ConsumeJMS	GetMongoRecord	PutHBaseRecord
ConsumeKafka	GetSFTP	PutHDFS
ConsumeKafka_0_10	GetSolr	PutHive3QL
ConsumeKafka_1_0	GetSplunk	PutHive3Streaming
ConsumeKafka_2_0	GetSQS	PutHiveQL
ConsumeKafka_2_6	GetTCP	PutHiveStreaming
ConsumeKafka2CDP	GetTwitter	PutHTMLElement
ConsumeKafka2RecordCDP	HandleHttpRequest	PutInfluxDB
ConsumeKafkaRecord_0_10	HandleHttpResponse	PutJMS
ConsumeKafkaRecord_1_0	HashAttribute	PutKafka
ConsumeKafkaRecord_2_0	HashContent	PutKinesisFirehose
ConsumeKafkaRecord_2_6	IdentifyMimeType	PutKinesisStream
ConsumeKinesisStream	InvokeAWSGatewayApi	PutKudu
ConsumeMQTT 1	InvokeGRPC	PutLambda
ConsumeWindowsEventLog	InvokeHTTP	PutMongoRecord
ControlRate	InvokeScriptedProcessor	PutORC
ConvertAvroSchema	JoltTransformJSON	PutParquet
ConvertAvroToJSON	JoltTransformRecord	PutRecord
ConvertAvroToORC	JsonQueryElasticsearch	PutRiemann
ConvertAvroToParquet	ListAzureBlobStorage	PutS3Object
ConvertCharacterSet	ListAzureDataLakeStorage	PutSFTP
ConvertCSVToAvro	ListCDPObjectStore	PutSNS
ConvertJSONToAvro	ListDatabaseTables	PutSolrContentStream
ConvertJSONToSQL	ListenFTP	PutSolrRecord
ConvertRecord	ListenGRPC	PutSplunk
CreateHadoopSequenceFile	ListenHTTP	PutSplunkHTTP 1
CryptographicHashAttribute	ListenRELp	PutSQL
CryptographicHashContent	ListenSyslog	PutSQS
DecryptContentPGP	ListenTCP	PutSyslog
DeleteAzureBlobStorage	ListenTCPRecord	PutTCP
DeleteAzureDataLakeStorage	ListenUDP	PutUDP
DeleteByQueryElasticsearch	ListenUDPRecord	PutWebSocket 1
DeleteCDPObjectStore	ListenWebSocket	QueryCassandra
DeleteDynamoDB	ListFile	QueryDatabaseTable
DeleteGCSObject	ListFTP	QueryDatabaseTableRecord
DeleteGridFS	ListGCSBucket	QueryElasticsearchHttp
DeleteHBaseCells	ListHDFS	QueryRecord
DeleteHBaseRow	ListS3	QuerySolr
DeleteHDFS	ListSFTP	QuerySplunkIndexingStatus
DeleteS3Object	LogAttribute	QueryWhois

DeleteSQS	LogMessage	ReplaceText 1
DetectDuplicate	LookupAttribute	ReplaceTextWithMapping
DistributeLoad	LookupRecord	ResizeImage
DuplicateFlowFile	MergeContent 1	RetryFlowFile
EncryptContent 2	MergeRecord	RouteHL7
EncryptContentPGP	ModifyHTMLElement	RouteOnAttribute
EnforceOrder	MonitorActivity	RouteOnContent
EvaluateJsonPath	Notify	RouteText
EvaluateXPath	PaginatedJsonQueryElasticsearch	SampleRecord 1
EvaluateXQuery	ParseCEF	ScanAccumulo
ExecuteGroovyScript	ParseEvtx	ScanAttribute
ExecuteInfluxDBQuery	ParseSyslog	ScanContent
ExecuteProcess	PartitionRecord	ScanHBase
ExecuteScript	PostHTTP	ScriptedFilterRecord
ExecuteSQL	PublishAMQP 1	ScriptedPartitionRecord
ExecuteSQLRecord	PublishGCPubSub 1	ScriptedTransformRecord
ExecuteStateless	PublishJMS 1	ScriptedValidateRecord
ExecuteStreamCommand	PublishKafka	ScrollElasticsearchHttp
ExtractAvroMetadata	PublishKafka_0_10	SearchElasticsearch
ExtractGrok	PublishKafka_1_0	SegmentContent
ExtractHL7Attributes	PublishKafka_2_0	SelectHive3QL
ExtractImageMetadata	PublishKafka_2_6	SelectHiveQL
ExtractText	PublishKafka2CDP	SignContentPGP
FetchAzureBlobStorage	PublishKafka2RecordCDP	SplitAvro 1
FetchAzureDataLakeStorage	PublishKafkaRecord_0_10	SplitContent 1
FetchCDPObjectStore	PublishKafkaRecord_1_0	SplitJson 1
FetchDistributedMapCache	PublishKafkaRecord_2_0	SplitRecord
FetchElasticsearch	PublishKafkaRecord_2_6	SplitText 1
FetchElasticsearchHttp	PublishMQTT 1	SplitXml 1
FetchFile	PutAccumuloRecord	TagS3Object
FetchFTP	PutAzureBlobStorage	TailFile
FetchGCSObject	PutAzureCosmosDBRecord 1	TransformXml
FetchGridFS	PutAzureDataLakeStorage	UnpackContent
FetchHBaseRow	PutAzureEventHub 1	UpdateAttribute
FetchHDFS	PutAzureQueueStorage	UpdateByQueryElasticsearch
FetchParquet	PutBigQueryBatch	UpdateCounter
FetchS3Object	PutBigQueryStreaming 1	UpdateHive3Table
FetchSFTP	PutCassandraQL 1	UpdateHiveTable
FlattenJson	PutCassandraRecord	UpdateRecord
ForkRecord	PutCDPObjectStore	ValidateCsv
GenerateFlowFile	PutCloudWatchMetric	ValidateRecord
GenerateTableFetch	PutCouchbaseKey 1	ValidateXml
GeoEnrichIP	PutDatabaseRecord	VerifyContentPGP
GeoEnrichIPRecord	PutDistributedMapCache	Wait
GetAzureEventHub	PutDynamoDB 1	YandexTranslate
GetAzureQueueStorage		
GetCouchbaseKey 1		
GetFile		



- 1 – indicates a memory intensive processor
- 2 – indicates a CPU intensive processor

## Supported NiFi Controller Services

This release ships with Apache NiFi 1.15.2 and includes a set of Controller Services, most of which are supported by Cloudera Support. You should be familiar with the available supported Controller Services, and avoid using any unsupported Controller Services in production environments.

Additional Controller Services are developed and tested by the Cloudera community but are not officially supported by Cloudera. Controller Services are excluded for a variety of reasons, including insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, and feature deviation from Cloudera best practices.

AccumuloService	HortonworksSchemaRegistry
ActionHandlerLookup	IPFIXReader
ADLSCredentialsControllerService	IPLookupService
ADLSIDBrokerCloudCredentialsProviderControllerService	JASN1Reader
AlertHandler	JMSConnectionFactoryProvider
AvroReader	JndiJmsConnectionFactoryProvider
AvroRecordSetWriter	JsonPathReader
AvroSchemaRegistry	JsonRecordSetWriter
AWSCredentialsProviderControllerService	JsonTreeReader
AWSIDBrokerCloudCredentialsProviderControllerService	KafkaRecordSink_1_0
AzureBlobIDBrokerCloudCredentialsProviderControllerService	KafkaRecordSink_2_0
AzureCosmosDBClientService	KafkaRecordSink_2_6
AzureStorageCredentialsControllerService	KerberosKeytabUserService
AzureStorageCredentialsControllerServiceLookup	KerberosPasswordUserService
CassandraDistributedMapCache	KerberosTicketCacheUserService
CassandraSessionProvider	KeytabCredentialsService
CouchbaseClusterService	KuduLookupService
CouchbaseKeyValueLookupService	LoggingRecordSink
CouchbaseMapCacheClient	LogHandler
CouchbaseRecordLookupService	MongoDBControllerService
CSVReader	MongoDBLookupService
CSVRecordLookupService	ParquetReader
CSVRecordSetWriter	ParquetRecordSetWriter
DatabaseRecordLookupService	PrometheusRecordSink
DatabaseRecordSink	ReaderLookup
DBCPCConnectionPool	RecordSetWriterLookup
DBCPCConnectionPoolLookup	RecordSinkHandler
DistributedMapCacheClientService	RecordSinkServiceLookup
DistributedMapCacheLookupService	RedisConnectionPoolService
DistributedMapCacheServer	RedisDistributedMapCacheClientService
DistributedSetCacheClientService	RestLookupService
DistributedSetCacheServer	ScriptedActionHandler
EasyRulesEngineProvider	ScriptedLookupService
EasyRulesEngineService	ScriptedReader
ElasticSearchClientServiceImpl	ScriptedRecordSetWriter
ElasticSearchLookupService	ScriptedRecordSink
ElasticSearchStringLookupService	ScriptedRulesEngine

EmailRecordSink	SimpleDatabaseLookupService
EmbeddedHazelcastCacheManager	SimpleKeyValueLookupService
ExpressionHandler	SimpleScriptedLookupService
ExternalHazelcastCacheManager	SiteToSiteReportingRecordSink
FreeFormTextRecordSetWriter	StandardHttpContextMap
GCPCredentialsControllerService	StandardPGPPrivateKeyService
GrokReader	StandardPGPPublicKeyService
HadoopDBCPCConnectionPool	StandardProxyConfigurationService
HazelcastMapCacheClient	StandardRestrictedSSLContextService
HBase_1_1_2_ClientMapCacheService	StandardS3EncryptionService
HBase_1_1_2_ClientService	StandardSSLContextService
HBase_1_1_2_ListLookupService	Syslog5424Reader
HBase_1_1_2_RecordLookupService	SyslogReader
HBase_2_ClientMapCacheService	VolatileSchemaCache
HBase_2_ClientService	WindowsEventLogReader
HBase_2_RecordLookupService	XMLReader
Hive3ConnectionPool	XMLRecordSetWriter
HiveConnectionPool	

## Supported NiFi Reporting Tasks

This release ships with Apache NiFi 1.15.2 and includes a set of Reporting Tasks, most of which are supported by Cloudera Support. You should be familiar with the available supported Reporting Tasks, and avoid using any unsupported Reporting Tasks in production environments.

Additional Reporting Tasks are developed and tested by the Cloudera community but are not officially supported by Cloudera. Reporting Tasks are excluded for a variety of reasons, including insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, and feature deviation from Cloudera best practices. Do not use these features in your production environments.

- AmbariReportingTask
- ControllerStatusReportingTask
- MetricsEventReportingTask
- MonitorDiskUsage
- MonitorMemory
- PrometheusReportingTask
- QueryNiFiReportingTask
- ReportLineageToAtlas
- ScriptedReportingTask
- SiteToSiteBulletinReportingTask
- SiteToSiteMetricsReportingTask
- SiteToSiteProvenanceReportingTask
- SiteToSiteStatusReportingTask

## Components Supported by Partners

This release ships with Apache NiFi 1.15.2 and includes a set of components built, maintained and supported by Cloudera partners. You should reach out directly to these partners in case you need assistance.

These components are not officially supported by Cloudera Support even though Cloudera Quality Engineering teams added test coverage for these components.

**Processors supported by partners**

- ConsumePulsar (v1.14.0-rc-3)
- ConsumePulsarRecord (v1.14.0-rc-3)
- PublishPulsar (v1.14.0-rc-3)
- PublishPulsarRecord (v1.14.0-rc-3)

**Controller Services supported by partners**

- PulsarClientAthenzAuthenticationService (v1.14.0-rc-3)
- PulsarClientJwtAuthenticationService (v1.14.0-rc-3)
- PulsarClientOAuthAuthenticationService (v1.14.0-rc-3)
- PulsarClientTlsAuthenticationService (v1.14.0-rc-3)
- StandardPulsarClientService (v1.14.0-rc-3)

These components can be used to push data into Apache Pulsar as well as getting data out of it. In case you have issues or questions while using these components, Cloudera recommends you to reach out to your StreamNative representative team.

## Unsupported Features in Cloudera DataFlow for Data Hub 7.2.14

Some features exist within Cloudera DataFlow for Data Hub 7.2.14 components, but are not supported by Cloudera.

### Unsupported Flow Management features

There are no unsupported Flow Management features in Cloudera DataFlow for Data Hub 7.2.14

#### NiFi

There are no updates for this release.

#### NiFi Registry

There are no updates for this release.

#### Related Information

[Cloudera Community Forum](#)

### Unsupported Streams Messaging features

Some Streams Messaging features exist in Cloudera DataFlow for Data Hub 7.2.14, but are not supported by Cloudera.

#### Kafka

The following Kafka features are not ready for production deployment. Cloudera encourages you to explore these features in non-production environments and provide feedback on your experiences through the *Cloudera Community Forums*.

- Only Java and .Net based clients are supported. Clients developed with C, C++, Python, and other languages are currently not supported.

- While Kafka Connect is available as part of Runtime, it is considered technical preview and is currently not supported in CDP Public Cloud. NiFi is a proven solution for batch and real time data loading that complement Kafka's message broker capability. For more information, see [Creating your first Flow Management cluster](#).
- The Kafka default authorizer is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.
- SASL/SCRAM is only supported for delegation token based authentication. It is not supported as a standalone authentication mechanism.

### Schema Registry

There are no updates for this release.

### Streams Messaging Manager

There are no updates for this release.

### Streams Replication Manager

There are no updates for this release.

### Cruise Control

There are no updates for this release.

### Related Information

[Cloudera Community Forum](#)

[Creating your first Streams Messaging cluster](#)

## Unsupported Streaming Analytics features

Some Streaming Analytic features exist in Cloudera DataFlow for Data Hub 7.2.14, but are not supported by Cloudera.

The following features are not ready for production deployment. Cloudera encourages you to explore these features in non-production environments and provide feedback on your experiences through the *Cloudera Community Forums*.

### Flink

- Apache Flink batch (DataSet) API
- GPU Resource Plugin
- Application Mode deployment
- SQL Client
- Python API
- The following features are not supported in SQL and Table API:
  - HBase Table Connector
  - Old Planner
  - Non-windowed (unbounded) joins, distinct

### Related Information

[Cloudera Community Forum](#)

## Known Issues In Cloudera DataFlow for Data Hub 7.2.14

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera DataFlow for Data Hub 7.2.14.

### Known Issues in Flow Management

Learn about the known issues in Flow Management clusters, the impact or changes to the functionality, and the workaround.

Learn about the known issues and limitations in Flow Management in this release:

#### **KafkaRecordSink puts multiple records in one message**

All the records are sent as a single Kafka message containing an array of records.

For more information, see [NIFI-8326](#).

There is no workaround for this issue.

#### **Kudu client preventing the creation of new tables using NiFi processors (KUDU-3297)**

There is an issue in the Kudu client preventing the creation of new tables using NiFi processors. The table needs to exist before NiFi tries to push data into it. You may see this error when this issue arises:

```
Caused by: org.apache.kudu.client.NonRecoverableException: failed to wait for Hive Metastore notification log listener to catch up: failed to retrieve notification log events: failed to open Hive Metastore connection: SASL(-15): mechanism too weak for this user
```

There is no workaround for this issue.

#### **NiFi Atlas reporting task does not work after data lake upgrade from light to medium**

After you upgrade your data lake from light to medium scale, the data lake machine hostname and IP address will change. As the Atlas reporting task uses Atlas and Kafka server hostnames, after the upgrade the wrong hostnames will prevent NiFi to report into Atlas.

Update the configuration of the ReportLineageToAtlas reporting task:

1. Open the Global menu on the NiFi UI.
2. Click Controller settings.
3. Select the Reporting tasks tab in the dialog box.
4. Stop the ReportLineageToAtlas reporting task and update the configuration:
  - Replace the hostname value in the Atlas Urls configuration with the new Atlas hostname.
  - Replace the hostnames value in the Kafka Bootstrap servers configuration with the new Kafka bootstrap server hostnames.
5. Start the ReportLineageToAtlas reporting task.

### Technical Service Bulletins

#### **TSB 2022-580: NiFi Processors cannot write to content repository**

If the content repository disk is filled more than 50% (or any other value that is set in `nifi.properties` for `nifi.content.repository.archive.max.usage.percentage`), and if there is no data in the content repository archive, the following warning message can be found in the logs: "Unable to write flowfile content to content repository container default due to archive file size constraints; waiting for archive cleanup". This would block the processors and no more data is processed.

This appears to only happen if there is already data in the content repository on startup that needs to be archived, or if the following message is logged: “Found unknown file XYZ in the File System Repository; archiving file”.

#### Upstream JIRA

- [NIFI-10023](#)
- [NIFI-9993](#)

#### Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2022-580: NiFi Processors cannot write to content repository](#)

#### TSB 2022-589: CVE-2022-33140 Apache NiFi ShellUserGroupProvider Vulnerability

The optional ShellUserGroupProvider in Apache NiFi 1.10.0 to 1.16.2 and Apache NiFi Registry 0.6.0 to 1.16.2 does not neutralize arguments for group resolution commands, allowing injection of operating system commands on Linux and macOS platforms. The ShellUserGroupProvider is not included in the default configuration. Command injection requires ShellUserGroupProvider to be one of the enabled User Group Providers (UGP) in the Authorizers configuration. Command injection also requires an authenticated user with elevated privileges. Apache NiFi requires an authenticated user with authorization to modify access policies in order to execute the command. Apache NiFi Registry requires an authenticated user with authorization to read user groups in order to execute the command. The resolution removes command formatting based on user-provided arguments.

#### Knowledge article

For the latest update on this issue see the corresponding Knowledge article: [TSB 2022-589: CVE-2022-33140 Apache NiFi ShellUserGroupProvider Vulnerability](#)

## Known Issues in Streams Messaging

Learn about the known issues in Streams Messaging clusters, the impact or changes to the functionality, and the workaround.

### Kafka

Learn about the known issues and limitations in Kafka in this release:

#### 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.

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-reassign-partitions`

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.

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.

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

#### **KAFKA-2561: 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.

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

#### **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.

None

#### **CDPD-49304: AvroConverter does not support composite default values**

AvroConverter cannot handle schemas containing a `STRUCT` type default value.

None.

#### **Limitations**

#### **Collection of Partition Level Metrics May Cause Cloudera Manager's Performance to Degrade**

If the Kafka service operates with a large number of partitions, collection of partition level metrics may cause Cloudera Manager's performance to degrade.

If you are observing performance degradation and your cluster is operating with a high number of partitions, you can choose to disable the collection of partition level metrics.



**Important:** If you are using SMM to monitor Kafka or Cruise Control for rebalancing Kafka partitions, be aware that both SMM and Cruise Control rely on partition level metrics. If partition level metric collection is disabled, SMM will not be able to display information about partitions. In addition, Cruise Control will not operate properly.

Complete the following steps to turn off the collection of partition level metrics:

1. Obtain the Kafka service name:
  - a. In Cloudera Manager, Select the Kafka service.
  - b. Select any available chart, and select Open in Chart Builder from the configuration icon drop-down.
  - c. Find `$SERVICENAME=` near the top of the display.

The Kafka service name is the value of `$SERVICENAME`.

2. Turn off the collection of partition level metrics:
  - a. Go to Hosts Configuration.
  - b. Find and configure the Cloudera Manager Agent Monitoring Advanced Configuration Snippet (Safety Valve) configuration property.

Enter the following to turn off the collection of partition level metrics:

```
[KAFKA_SERVICE_NAME]_feature_send_broker_topic_partition_entity_update_enabled=false
```

Replace [KAFKA\_SERVICE\_NAME] with the service name of Kafka obtained in step 1. The service name should always be in lower case.

- c. Click Save Changes.

## Schema Registry

### CDPD-49304: AvroConverter does not support composite default values

AvroConverter cannot handle schemas containing a STRUCT type default value.

None.

### CDPD-54379: KafkaJsonSerializer and KafkaJsonDeserializer do not allow null values

KafkaJsonSerializer and KafkaJsonDeserializer do not allow the data to be null, resulting in a NullPointerException (NPE).

None.

### CDPD-49217 and CDPD-50309: Schema Registry caches user group membership indefinitely

Schema Registry caches the Kerberos user and group information indefinitely and does not catch up on group membership changes.

Restart Schema Registry after group membership changes.

### CDPD-60160: Schema Registry Atlas integration does not work with Oracle databases

Schema Registry is unable to create entities in Atlas if Schema Registry uses an Oracle database. The following will be present in the Schema Registry log if you are affected by this issue:

```
ERROR com.cloudera.dim.atlas.events.AtlasEventsProcessor: An error occurred while processing Atlas events.
java.lang.IllegalArgumentException: Cannot invoke com.hortonworks.registries.schemaregistry.AtlasEventStorable.setType on bean class 'class com.hortonworks.registries.schemaregistry.AtlasEventStorable' - argument type mismatch - had objects of type "java.lang.Long" but expected signature "java.lang.Integer"
```

This issue causes the loss of audit data on Oracle environments.

None.

### CDPD-58949: Schemas are de-duplicated on import

On import, Schema Registry de-duplicates schema versions based on their fingerprints. This means that schemas which are considered functionally equivalent in SR get de-duplicated. As a result, some schema versions are not created, and their IDs do not become valid IDs in SR.

None.

### CDPD-58990: getSortedSchemaVersions method orders by schemaVersionId instead of version number

On validation, Schema Registry orders schema versions based on ID instead of version number. In some situations, this can cause validation with the LATEST level to compare the new schema version to a non-latest version.



This situation can occur when an older version of a schema has a higher ID than the newer version of a schema, for example, when the older version is imported with an explicit ID.

None.

## Streams Messaging Manager

Learn about the known issues in Streams Messaging Manager in this release.

### **CDPD-33770: On the topics details page selecting a custom timestamp is broken**

When you select a custom TimePeriod (a non-predefined TimePeriod like 6 hours, 30 minutes etc.) on the SMM UI's topicDetail page, an error is going to be thrown, and the replication related metrics would not be displayed.

### **CDPD-46728: SMM UI shows the consumerGroup instead of the instances on the Profile page's right hand side**

On the ConsumerGroupDetail page, SMM UI shows the group instead of its instances on the right hand side table.

None.

### **OPSAPS-63017: The Kafka Connect tab is missing from the SMM UI**

Under certain circumstances the Kafka Connect tab in SMM might not be available by default on Data Hub clusters even if Kafka Connect is provisioned on the cluster. As a result, interacting with Kafka Connect using SMM is not possible.

1. Access the Cloudera Manager instance managing the affected Data Hub cluster.
2. Select the Streams Messaging Manager service, and go to Configuration.
3. Find and configure the following properties:

- Kafka Connect Host

Enter the hostname of the machine that the Kafka Connect role is deployed on. If you have multiple instances of the Kafka Connect role, you can choose to use any of them. Add a single hostname, as configuring multiple hostnames for high availability is currently not supported.

- Kafka Connect Port

Enter the port that the Kafka Connect role is using. The value of this property must match the port set in the Secure Kafka Connect Rest Port Kafka property.

- Kafka Connect Protocol

Set this property to https.

### **OPSAPS-59553: SMM's bootstrap server config should be updated based on Kafka's listeners**

SMM does not show any metrics for Kafka or Kafka Connect when multiple listeners are set in Kafka.

SMM cannot identify multiple listeners and still points to bootstrap server using the default broker port (9093 for SASL\_SSL). You would have to override bootstrap server URL (hostname:port as set in the listeners for broker). Add the bootstrap server details in SMM safety valve in the following path:

Cloudera Manager SMM Configuration Streams Messaging Manager Rest Admin Server  
Advanced Configuration Snippet (Safety Valve) for streams-messaging-manager.yaml Add the following value for bootstrap servers Save Changes Restart SMM :

```
streams.messaging.manager.kafka.bootstrap.servers=<comma-separated list of brokers>
```

### **OPSAPS-59597: SMM UI logs are not supported by Cloudera Manager**

Cloudera Manager does not support the log type used by SMM UI.

View the SMM UI logs on the host.

## Streams Replication Manager

Learn about the known issues and limitations in Streams Replication Manager in this release:

### Known Issues

#### **CDPD-22089: SRM does not sync re-created source topics until the offsets have caught up with target topic**

Messages written to topics that were deleted and re-created are not replicated until the source topic reaches the same offset as the target topic. For example, if at the time of deletion and re-creation there are a 100 messages on the source and target clusters, new messages will only get replicated once the re-created source topic has 100 messages. This leads to messages being lost.

None

#### **CDPD-30275: SRM may automatically re-create deleted topics on target clusters**

If `auto.create.topics.enable` is enabled, deleted topics might get automatically re-created on target clusters. This is a timing issue. It only occurs if remote topics are deleted while the replication of the topic is still ongoing.

1. Remove the topic from the topic allowlist with `srm-control`. For example:

```
srm-control topics --source [SOURCE_CLUSTER] --target [TARGET_CLUSTER] --remove [TOPIC1]
```

2. Wait until SRM is no longer replicating the topic.
3. Delete the remote topic in the target cluster.

#### **OPSAPS-63104: The automatically generated password for co-located services is invalid**

SRM automatically generates a username and password that can be used by co-located services to access SRM and its REST API. However, a unique password is generated for each SRM Service role instance. Because of this, co-located services that use the password, for example SMM, can only connect to one of the SRM Service role instances.

Manually configure a password using the SRM Service Co-Located Service User Password SRM property. The password you configure will be accepted by all SRM Service role instances.

#### **CDPD-11079: Blacklisted topics appear in the list of replicated topics**

If a topic was originally replicated but was later disallowed (blacklisted), it will still appear as a replicated topic under the `/remote-topics` REST API endpoint. As a result, if a call is made to this endpoint, the disallowed topic will be included in the response. Additionally, the disallowed topic will also be visible in the SMM UI. However, its Partitions and Consumer Groups will be 0, its Throughput, Replication Latency and Checkpoint Latency will show N/A.

None

#### **OPSAPS-62546: Kafka External Account SSL keypassword configuration is used incorrectly by SRM**

When a Kafka External Account specifies a keystore that uses an SSL key password, SRM uses it as the `ssl.keystore.key` configuration. Due to using the incorrect `ssl.keystore.key` configuration, SRM will fail to load the keystore in certain cases.

Workaround: For the keystores used by the Kafka External Accounts, the SSL key password should match the SSL keystore password, and the SSL keystore key password should not be provided. Alternatively, you can use the legacy connection configurations based on the `streams.replication.manager.configs` to specify the SSL key password.

### Limitations

#### **SRM cannot replicate Ranger authorization policies to or from Kafka clusters**

Due to a limitation in the Kafka-Ranger plugin, SRM cannot replicate Ranger policies to or from clusters that are configured to use Ranger for authorization. If you are using SRM to replicate data

to or from a cluster that uses Ranger, disable authorization policy synchronization in SRM. This can be achieved by clearing the Sync Topic Acls Enabled (sync.topic.acls.enabled) checkbox.

### SRM cannot ensure the exactly-once semantics of transactional source topics

SRM data replication uses at-least-once guarantees, and as a result cannot ensure the exactly-once semantics (EOS) of transactional topics in the backup/target cluster.



**Note:** Even though EOS is not guaranteed, you can still replicate the data of a transactional source, but you must set isolation.level to read\_committed for SRM's internal consumers. This can be done by adding `[***CONFIG LEVEL PREFIX***].isolation.level=read_committed` to the Streams Replication Manager's Replication Configs SRM service property in Cloudera Manager. The isolation.level property can be set on a global connector or replication level. For example:

```
#Global connector level
connectors.consumer.isolation.level=read_committed
#Replication level
uswest->useast.consumer.isolation.level=read_committed
```

### SRM checkpointing is not supported for transactional source topics

SRM does not correctly translate checkpoints (committed consumer group offsets) for transactional topics. Checkpointing assumes that the offset mapping function is always increasing, but with transactional source topics this is violated. Transactional topics have control messages in them, which take up an offset in the log, but they are never returned on the consumer API. This causes the mappings to decrease, causing issues in the checkpointing feature. As a result of this limitation, consumer failover operations for transactional topics is not possible.

## Cruise Control

Learn about the known issues and limitations in Cruise Control in this release:

### CDPD-47616: Unable to initiate rebalance, number of valid windows (NumValidWindows) is zero

If a Cruise Control rebalance is initiated with the rebalance\_disk parameter and Cruise Control is configured to fetch metrics from Cloudera Manager (Metric Reporter is set to CM metrics reporter), Cruise Control stops collecting metrics from the partitions that are moved. This is because Cloudera Manager does not collect metrics from moved partitions due to an issue in Kafka (KAFKA-10320).

If the metrics are not available, the partition is considered invalid by Cruise Control. This results in Cruise Control blocking rebalance operations and proposal generation.

Configure Cruise Control to use the Cruise Control metrics reporter (default). This issue is not present if this metric reporter is used.

1. In Cloudera Manager, select the Cruise Control service.
2. Go to Configuration.
3. Find the Metric Reporter property.
4. Select the Cruise Control metrics reporter option.
5. Restart the Cruise Control service.

### OPSAPS-68148: Cruise Control rack aware goal upgrade handler

The goal sets in Cruise Control, which include the default, supported, hard, self-healing and anomaly detection goals, might be overridden to their default value after a cluster upgrade if the goals have been customized.

Create a copy from the values of the goal lists before upgrading your cluster, and add the copied values to the goal lists after upgrading the cluster. Furthermore, you must rename any mentioning of `com.linkedin.kafka.cruisecontrol.analyzer.goals.RackAwareGoal` to `com.linkedin.kafka.cruisecontrol.analyzer.goals.RackAwareDistributionGoal` as Cruise Control will not be able to start otherwise.

## Known Issues in Streaming Analytics

Learn about the known issues in Streaming Analytics clusters, the impact or changes to the functionality, and the workaround.

### SQL Stream Builder

#### **FLINK-18027: ROW value constructor cannot deal with complex expressions**

When querying data from a table or a view with a ROW() function an exception is thrown due to a Calcite parsing issue. For example, the following query will return an error:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM table;
SELECT * FROM example;
```

Add a second SELECT layer to the SQL query as shown in the following example:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM (SELECT col1,
col2 FROM table);
SELECT * FROM example;
```

#### **Uploading connector files fail**

When trying to upload a new connector JAR with a size file more than 1 MB, the upload process fails with an error.

Set the server.tomcat.max-swallow-size in Cloudera Manager using the following steps:

1. Open your cluster in Cloudera Manager.
2. Select SQL Stream Builder from the list of services.
3. Select Configuration.
4. Search for Streaming SQL Engine Advanced Configuration Snippet (Safety Valve) for ssb-conf/application.properties in the search bar.
5. Add server.tomcat.max-swallow-size=2000MB to the **Safety Valve**.
6. Click Save.
7. Restart the SQL Stream Builder service.

#### **CSA-3742: Catalogs are not working due to expired Kerberos TGT**

When SSB is running for a longer period of time than the lifetime of the Kerberos Ticket Granting Ticket (TGT), authentication with the catalog services will fail and the catalogs stop working.

None

#### **CSA-2016: Deleting table from other teams**

There is a limitation when using the Streaming SQL Console for deleting tables. It is not possible to delete a table that belongs to another team using the Delete button on the User Interface.

Use DROP TABLE statement from the SQL window.

#### **CSA-1454: Timezone settings can cause unexpected behavior in Kafka tables**

You must consider the timezone settings of your environment when using timestamps in a Kafka table as it can affect the results of your query. When the timestamp in a query is identified with from\_unixtime, it returns the results based on the timezone of the system. If the timezone is not set in UTC+0, the timestamp of the query results will shift in time and will not be correct.

Change your local timezone settings to UTC+0.

#### **CSA-1232: Big numbers are incorrectly represented on the Streaming SQL Console UI**

The issue impacts the following scenarios in Streaming SQL Console:

- When having integers bigger than 253-1 among your values, the Input transformations and User Defined Functions are considered unsafe and produce incorrect results as these numbers will lose precision during parsing.
- When having integers bigger than 253-1 among your values, sampling to the Streaming SQL Console UI produces incorrect results as these numbers will lose precision during parsing.

None

## Flink

### FLINK-18027: ROW value constructor cannot deal with complex expressions

When querying data from a table or a view with a ROW() function an exception is thrown due to a Calcite parsing issue. For example, the following query will return an error:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM table;
SELECT * FROM example;
```

Add a second SELECT layer to the SQL query as shown in the following example:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM (SELECT col1,
col2 FROM table);
SELECT * FROM example;
```

- Match recognize
- Top-N
- Stream-Table join (without rowtime input)

### DataStream conversion limitations

- Converting between Tables and POJO DataStreams is currently not supported in CSA.
- Object arrays are not supported for Tuple conversion.
- The java.time class conversions for Tuple DataStreams are only supported by using explicit TypeInformation: LegacyInstantTypeInfo, LocalTimeTypeInfo.getInfoFor(LocalDate/LocalDateTime/LocalTime.class).
- Only java.sql.Timestamp is supported for rowtime conversion, java.time.LocalDateTime is not supported.

### Kudu catalog limitations

- CREATE TABLE
  - Primary keys can only be set by the kudu.primary-key-columns property. Using the PRIMARY KEY constraint is not yet possible.
  - Range partitioning is not supported.
- When getting a table through the catalog, NOT NULL and PRIMARY KEY constraints are ignored. All columns are described as being nullable, and not being primary keys.
- Kudu tables cannot be altered through the catalog other than simply renaming them.

### Schema Registry catalog limitations

- Currently, the Schema Registry catalog / format only supports reading messages with the latest enabled schema for any given Kafka topic at the time when the SQL query was compiled.
- No time-column and watermark support for Registry tables.
- No CREATE TABLE support. Schemas have to be registered directly in the SchemaRegistry to be accessible through the catalog.
- The catalog is read-only. It does not support table deletions or modifications.
- By default, it is assumed that Kafka message values contain the schema id as a prefix, because this is the default behaviour for the SchemaRegistry Kafka producer format. To consume messages with schema written in the header, the following property must be set for the Registry client: store.schema.version.id.in.header: true.

## Fixed Issues in Cloudera DataFlow for Data Hub 7.2.14

Fixed issues represent selected issues that were previously logged through Cloudera Support, but are addressed in the current release. These issues may have been reported in previous versions within the Known Issues section; meaning they were reported by customers or identified by Cloudera Quality Engineering team.

Review the list of issues that are resolved in Cloudera DataFlow for Data Hub 7.2.14.

### Fixed Issues in Flow Management

Review the list of Flow Management issues that are resolved in Cloudera DataFlow for Data Hub 7.2.14.

#### 7.2.14.3

##### Technical Service Bulletins

##### **TSB 2022-580: NiFi Processors cannot write to content repository**

For the latest update on this issue see the corresponding Knowledge article: [TSB 2022-580: NiFi Processors cannot write to content repository](#)

##### **TSB 2022-589: CVE-2022-33140 Apache NiFi ShellUserGroupProvider Vulnerability**

For the latest update on this issue see the corresponding Knowledge article: [TSB 2022-589: CVE-2022-33140 Apache NiFi ShellUserGroupProvider Vulnerability](#)

#### 7.2.14

##### **NIFI-9552**

Make sure cl-over-slf4j is included under ext/ranger/install/lib directory.

##### **NIFI-9534**

Upgraded Log4j 2 BOM from 2.17.0 to 2.17.1.

##### **NIFI-9524**

Exclude commons-logging and log4j-core banned dependencies for other build profiles as well.

##### **NIFI-9394**

Removed RequestLogger and TimerFilter.

##### **NIFI-9385**

Add Flow Metrics producer for Cloudera Manager.

##### **NIFI-9384**

Corrected usage and generics in ListenTCP.

##### **NIFI-9382**

Improve startup time when loading flow that uses many HDFS related processors.

##### **NIFI-9379**

Add dependent properties and resource definitions to manifest model.

##### **NIFI-9371**

Removed synchronized keyword from Active Threads methods.

##### **NIFI-9362**

Ensure that we update the StateMap in AbstractListProcessor to hold any files whose date matches the latest before setting cluster-wide state.

##### **NIFI-9360**

Update PutHDFS to handle filesystems which do not support getAclStatus().

**NIFI-9355**

Upgraded Apache Curator from 4.2.0 to 5.2.0.

**NIFI-9339**

Fixed headers to include appropriate Request-Token header for all requests to the server for JoltTransformJSON UI.

**NIFI-9338**

Add Azure Blob processors using Azure Blob Storage client library v12 for Java.

**NIFI-9335**

Updated AvroTypeUtil#createAvroRecord to ensure that if the given Avro Schema contains a field whose value is defaulted, the produced Avro Record has that value populated.

**NIFI-9334**

Add support for upsert in 'PutMongoRecord'. Use 'bulkWrite' for both insert and upsert.

**NIFI-9329**

Expose event validation in ParseCEF processor.

**NIFI-9328**

Transfer cleanup and reuse added to FetchFileTransfer in case of FileNotFound and PermissionDenied exceptions.

**NIFI-9311**

When determining property values, be sure to fetch the property descriptor from the component itself, rather than using the PropertyDescriptor in the Map. This allows us to ensure that if the definition of the PropertyDescriptor changes, the most up-to-date definition is picked up.

**NIFI-9308**

Added EmailRecordSink.

**NIFI-9304**

Added Google Cloud Pub/Sub Lite processors.

**NIFI-9300**

Fix AWSCredentialsService EL attribute evaluation.

**NIFI-9291**

Added NiFi HTTP request logging.

**NIFI-9289**

On startup, when enabling a Controller Service & its dependencies, do not wait for the dependencies to fully enable. Doing so can take 30 seconds per each Controller Service (and per each reference). Due to some previous refactoring, this waiting period is no longer necessary, as the referencing service can now be enabled and will asynchronously complete the enabling once it becomes valid (due to the referenced service becoming enabled).

**NIFI-9277**

Add Record Reader and Writer to ListenHTTP.

**NIFI-9265**

Fixing path handling for HDFS processors when there are multiplied separators in the path.

**NIFI-9260**

Making the 'write and rename' behaviour optional for PutHDFS.

**NIFI-9241**

Refactored CSRF mitigation using random Request-Token.

**NIFI-9235**

Log conflicts between umask and ACL in PutHDFS.

**NIFI-9231**

Add support for custom extensions in ParseCEF.

**NIFI-9229**

Flow upgrade not possible if a Output Port changes to a funnel.

**NIFI-9228**

Refactored tests using TemporaryKeyStoreBuilder.

**NIFI-9228**

Refactored tests to use generated KeyStores.

**NIFI-9217**

Avoid deadlock on cluster operation.

**NIFI-9210**

Upgraded jsoup from 1.8.3 to 1.14.2 - custom fix without minifi changes.

**NIFI-9205**

Update prioritizer configuration.

**NIFI-9202**

Improve Allowable Values merging to handle cases when different nodes have different set of Allowable Values.

**NIFI-9201**

NullPointerException in AbstractKerberosUser if the tgt is not renewable.

**NIFI-9200**

Free cache on heap after disabling AbstractCSVLookupService.

**NIFI-9192**

ResultSetRecordSet considers value of useLogicalTypes flag when determining the object's schema.

**NIFI-9183**

Add a command-line option to save status history.

**NIFI-9182**

When calling ProcessSession.append(), do not allow calls from processor to flush the underlying BufferedOutputStream. Instead, wrap in a NonFlushableOutputStream and only flush when session commit is called.

**NIFI-9148**

Refactored nifi-scripting-bundle to use JUnit 5.

**NIFI-9147**

Refactored nifi-rules-action-handler-bundle to use JUnit 5.

**NIFI-9146**

Refactored nifi-riemann-bundle to use JUnit 5.

**NIFI-9144**

Refactored nifi-registry-bundle to use JUnit 5.

**NIFI-9140**

Refactored nifi-prometheus-bundle using JUnit 5.

**NIFI-9139**

Refactored nifi-poi-bundle using JUnit 5.



**NIFI-9138**

Refactored nifi-pgp-bundle using JUnit 5.

**NIFI-9137**

Refactored nifi-parquet-processors using JUnit 5.

**NIFI-9082**

Add nifi.zookeeper.jute.maxbuffer property.

**NIFI-9079**

Set log level to WARN for Apache Atlas client logs.

**NIFI-9076**

HDFS operations in MoveHDFS wrapped in UGI.doAs().

**NIFI-9066**

Supporting flow file attributes in PutSplunkHTTP; Fixing endpoint assembly with extra query values.

**NIFI-9061**

Eliminated the nifi.cluster.node.protocol.threads property in favor of nifi.cluster.node.protocol.max.threads property so that we can properly scale out the number of threads used for HTTP request replication. Implementing a caching mechanism for creating the DateTimeFormatter used by TimeAdapter in order to improve performance when parsing timestamps in web requests. Implementing caching logic for caching the number of characters that can rendered without needing an ellipsis for some components in the UI.

**NIFI-9060**

Refactored HTTP Cookie Path Handling.

**NIFI-9055**

Added handling for 0- read range to FetchS3Object.

**NIFI-9054**

Calling Nifi Registry's createExtensionBundleVersion REST endpoint will cause NullPointerException.

**NIFI-9049**

Replaced localStorage with sessionStorage for Bearer Token.

**NIFI-9041**

Replaced JUnit 4 only testing configuration with a combination of JUnit 5 and JUnit Vintage.

**NIFI-9038**

Fix fingerprinting group access control policies for Remote Port.

**NIFI-9035**

Refactored isKeystoreValid() to avoid NullPointerException.

**NIFI-9032**

Refactoring HDFS processors in order to increase flexibility.

**NIFI-9025**

Fixed reference in TestTailFile unit test.

**NIFI-9018**

When connection points to a moved port version change with NiFi Registry may throw exception.

**NIFI-9017**

Update Load Balanced Connection logic so that if a node connects to the cluster with a different load balancing hostname/port, it starts sending to the new endpoint instead of failing to send to the old endpoint.

**NIFI-9009**

Components Validation backend work.

**NIFI-8996**

Close JDBC statements in PutHive\*QL processors.

**NIFI-8990**

Downgrade Gremlin from 3.5.1 to 3.4.4 to support Graph DBs.

**NIFI-8990**

Upgraded Groovy to 2.5.14.

**NIFI-8987**

Upgraded Tika to 1.27 and Graphics2d to 0.32.

**NIFI-8986**

Upgraded Commons Compress to 1.21 - custom fix without minifi changes.

**NIFI-8969**

Fix Maximum Polling Time in CuratorLeaderElectionManager.

**NIFI-8965**

Fix duplicate code and typo in StandardFlowManager.

**NIFI-8957**

NiFi Registry - Possibility to set a description when creating a bucket.

**NIFI-8955**

Add Max Connection Lifetime property to Hive(\_1\_1)ConnectionPool CS.

**NIFI-8951**

Fixed scale vs precision reference in QueryDatabaseTable processors.

**NIFI-8948**

Upgraded Spring to 5.3.9 and Security to 5.5.1.

**NIFI-8942**

NiFi Registry - flow description cannot be selected & copied on the UI.

**NIFI-8939**

Ensure that when async/long-running flow updates are made.

**NIFI-8938**

Ensure consistent version of jackson-core, jackson-databind, and jackson-annotations.

**NIFI-8937**

Show component name and version in configure dialog's title bar.

**NIFI-8936**

Added dynamic http header support to Confluent Schema Registry controller.

**NIFI-8933**

Configure Jersey's ObjectMapper to ignore unknown fields.

**NIFI-8931**

Removed OTP Authentication.

**NIFI-8928**

Upgrade Jetty to 9.4.43.v20210629.

**NIFI-8806**

Refactored ListenTCP using Netty.

**NIFI-8790**

Allow Expression Language for Index Operation in PutElasticsearchRecord.

**NIFI-8788**

Upgraded dependencies and removed unnecessary log4j test dependencies.

**NIFI-8787**

Wrapped hdfs.exists() call in UGI.doAs() in GetHDFS processor.

**NIFI-8785**

Confluent Schema Registry REST client refactoring.

**NIFI-8782**

Added Rate-Limiting for Access Token Requests.

**NIFI-8773**

Implemented Line Start Pattern in TailFile.

**NIFI-8770**

Use queue drainTo() on shutdown in HandleHttpRequest.

**NIFI-8768**

Added toLocalDate() for convertType() handling of DATE fields.

**NIFI-8766**

Implemented RS512 Algorithm for JWT Signing.

**NIFI-8762**

ADLSCredentialControllerService does not support EL for Storage Account name.

**NIFI-8761**

Enable not setting a value for Escape Character in CSVReader.

**NIFI-8759**

ExecuteSQL and ExecuteSQLRecord unnecessarily fall back to default decimal scale.

**NIFI-8749**

Removed implicit time zone conversion to GMT.

**NIFI-8739**

Penalized flowfiles should be able to be polled from the queue in some cases.

**NIFI-8727**

Addressed bug in which ProcessSession doesn't properly decrement claimant count when a FlowFile is cloned and then the clone written to. Added automated tests to ensure that we are properly handling cases where a FlowFile is clone and then the contents modified.

**NIFI-8708**

Upgraded Spring Framework to 5.3.8 for several extensions.

**NIFI-8668**

ConsumeAzureEventHub NiFi processors need to support storage SAS token authentication.

**NIFI-8639**

Add incoming flowfile to ConnectWebSocket processor to configure custom headers and dynamic URL through flowfile attributes in JettyWebSocketClient service.

**NIFI-8442**

Add a new test with date, timestamp and time as string & New management of date, time and timestamp NIFI-8442 Put DateTimeFormatter as static and Add comments to explain why ZoneOffset.UTC is required.

**NIFI-8439**

Update parquet-avro to allow reading parquet INT96 timestamps as byte arrays (instead of throwing an exception).

**NIFI-8385**

Add FlowFiles from logging to bulletins.

**NIFI-8376**

Gracefully handle SQL exceptions in ResultSetRecordSet.

**NIFI-8273**

Adding Scripted Record processors.

**NIFI-8240**

Unify Kudu versions.

**NIFI-7947**

Add directory deletion functionality in DeleteAzureDataLakeStorage.

**NIFI-7443**

Corrected SFTP Keep Alive behavior.

**NIFI-7012**

Refactored OnConfigurationRestored to support sensitive property validation.

**NIFI-5936**

Added DROP provenance event to MockProcessSession.remove() to match real impl.

**NIFI-4542**

Add target.dir.created to indicate if the target directory created.

**NIFI-3328**

SendTrapSNMP and ListenTrapSNMP processors added.

**NIFI-830**

Added FlowFile Naming Strategy to InvokeHTTP.

## Fixed Issues in Streams Messaging

Review the list of Streams Messaging issues that are resolved in Cloudera DataFlow for Data Hub 7.2.14.

**Kafka****CDPD-27780: IdentityReplicationPolicy for MM2 to mimic MM1 (KAFKA-9726 backport)**

**Warning:** The IdentityReplicationPolicy does not detect cycles. As a result, using this replication policy is only viable in deployments where the replication setup is acyclic. If your replication setup is not acyclic, using this replication policy might result in records being replicated in an infinite loop between clusters. Additionally, monitoring replications with the SRM Service is not possible when this policy is in use.

This is a backported improvement that introduces a new replication policy called IdentityReplicationPolicy. This replication policy does not rename remote (replicated) topics. As a result of this backport, the IdentityReplicationPolicy is now available for use with Streams Replication Manager. For more information, see [KAFKA-9726](#).

**CDPD-29307: Kafka producer entity stays in incomplete state in Atlas**

The Kafka-Atlas plugin now creates Producer and Consumer entities correctly.

**OPSAPS-61697: Kafka broker IDs are overridden when importing a cluster template**

Importing a cluster template to a new cluster in Cloudera Manager no longer overrides Kafka broker IDs (broker.id) or other role specific unique identifiers if the unique identifiers are already set in the template.

**Schema Registry****CDPD-32192: First start failed for SR, with oracle DB, migration failed at CREATE TABLE "atlas\_events"**

Fixed v009\_\_create\_registry\_audit.sql to have create index refer to the lower case "atlas\_events" object (the table).

Made the script rerunnable since the table was already created where the script had already run.

**CDPD-31907: Schema Registry REST API endpoint does not show SchemaBranches**

Schema Registry's /api/v1/schemaregistry/schemas/aggregated REST API endpoint shows SchemaBranches without SchemaVersions.

**CDPD-30996: SR does not create new SchemaMetadata with given ID**

In DefaultSchemaRegistry class, addSchemaMetadata (Supplier<Long> id, SchemaMetadata schemaMetadata, boolean throwErrorIfExists) does not look for the given ID, but the next available ID.

**CDPD-29700: Hide Compatibility list in the website**

When the schema type is JSON, then the Compatibility field will be hidden in the website.

**CDPD-29663: Error while connecting topic with schema in Atlas**

When Schema Registry tries to make a relationship in Atlas between a schema and a non-existent corresponding topic, an error occurs.

**CDPQE-11299: Importing schemas in Confluent format might fail**

Fixed the issue where importing from the Confluent Schema Registry fails intermittently.

**Streams Messaging Manager****CDPD-30745: Broken link to the TopicDetailPage**

When a topic is selected in the Replications tab, where the topic in question is not present on the local Kafka Cluster monitored by the current SMM instance, link breaks to the TopicDetailPage.

**CDPD-30731: Knox rewrite not happening for new v2 SMM endpoints**

SMM is not showing the Replications tab on Public Cloud when Knox is enabled.

**CDPD-30370: When TLS is enabled, SMM should connect to Schema Registry**

When TLS is enabled, SMM by default cannot properly connect to Schema Registry. As a result, the SMM Data Explorer shows errors when viewing Avro formatted data.

**CDPD-28002: The Cluster Replications tab is missing from the SMM UI**

SMM UI now correctly renders Replication workflows and metrics in Public Cloud.

**CDPD-24943: Long resource names (such as topic names, hostnames etc) are truncated on SMM UI**

Resource names in listings (such as topic name, host name, consumer name and so on ) are now overflowing to the next line with breaking text on any character, instead of remaining in 1 line with hidden overflow.

**Streams Replication Manager****OPSAPS-61814: Using the service dependency method to define Kerberos enabled co-located clusters is not supported**

When the Streams Replication Manager Co-located Kafka Cluster Alias configuration is used to auto-configure the connection to the co-located Kafka cluster, and Kerberos is enabled, the JAAS configuration is dynamically generated on each host. As a result, you can now use the service dependency method to define a Kerberos enabled co-located cluster.

**CDPD-31235: Negative consumer group lag when replicating groups through SRM**

SRM no longer tries to create a checkpoint or synchronize the group offset if there is no mapping available for the topic-partition in the offset-syncs topic.

**Cruise Control****CDPD-33535: Upgrading Logredactor version**

The Logredactor version is upgraded to 2.0.13 version to fix CVE-2021-44228 issues.

## Fixed Issues in Streaming Analytics

Review the list of Streaming Analytics issues that are resolved in Cloudera DataFlow for Data Hub 7.2.14.

**7.2.14.6****CSA-3742: Catalogs are not working due to expired Kerberos TGT**

The issue regarding the expired Kerberos Ticket Granting Ticket (TGT) and catalog authentication has been fixed.

**7.2.14****CSA-1318: Flink job submission fails in RAZ enabled environment**

Flink applications can be run in environments where Ranger Authorization Service (RAZ) is enabled.

**CSA-1985: DROP TABLE limitation when using Webhook table**

The issue of SQL Stream Builder (SSB) operations not showing in Atlas has been fixed. Atlas can be used to monitor SSB jobs.

**CSA-1673: SSB operations are not showing in Atlas**

The issue of SQL Stream Builder (SSB) operations not showing in Atlas has been fixed. Atlas can be used to monitor SSB jobs.