

Cloudera DataFlow for Data Hub Release Notes

Date published: 2019-12-16

Date modified: 2023-01-11



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

| | |
|---|---------------|
| What's New in Cloudera DataFlow for Data Hub 7.2.16..... | 4 |
| What's New in Flow Management..... | 4 |
| What's New in Streams Messaging..... | 4 |
| What's New in Streaming Analytics..... | 9 |
| Component Support in Cloudera DataFlow for Data Hub 7.2.16..... | 11 |
| Supported NiFi Extensions..... | 12 |
| Supported NiFi Processors..... | 12 |
| Supported NiFi Controller Services..... | 14 |
| Supported NiFi Reporting Tasks..... | 16 |
| Supported NiFi Parameter Providers..... | 16 |
| Components supported by partners..... | 17 |
| Unsupported Features in Cloudera DataFlow for Data Hub 7.2.16..... | 17 |
| Unsupported Flow Management features..... | 17 |
| Unsupported Streams Messaging features..... | 18 |
| Unsupported Streaming Analytics features..... | 19 |
| Known Issues In Cloudera DataFlow for Data Hub 7.2.16..... | 19 |
| Known Issues in Flow Management..... | 19 |
| Known Issues in Streams Messaging..... | 20 |
| Known Issues in Streaming Analytics..... | 28 |
| Fixed Issues in Cloudera DataFlow for Data Hub 7.2.16..... | 33 |
| Fixed Issues in Flow Management..... | 33 |
| Fixed Issues in Streams Messaging..... | 37 |
| Fixed Issues in Streaming Analytics..... | 40 |
| Fixed CVEs in Cloudera DataFlow for Data Hub 7.2.16..... | 41 |
| CVE-2021-45105 & CVE-2021-44832 remediation for CDF for Data Hub..... | 41 |
| Fixed CVEs in Flow Management..... | 42 |
| Behavioral Changes in Cloudera DataFlow for Data Hub 7.2.16..... | 42 |
| Behavioral Changes in Streams Messaging..... | 42 |
| Behavioral Changes in Streaming Analytics..... | 43 |

What's New in Cloudera DataFlow for Data Hub 7.2.16

Cloudera DataFlow for Data Hub 7.2.16 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 7.2.16.

Flow Management DataHub in CDP 7.2.16 is based on Apache NiFi 1.18.0 and includes significant improvements and fixes. Here are the most important new features and improvements:

- PutIceberg processor

This release includes in Technical Preview a PutIceberg processor to ingest data using Iceberg table format.

- Snowflake integration based on Snowpipe

This release includes in Technical Preview a set of processors (GetSnowflakeIngestStatus, StartSnowflakeIngest, and PutSnowflakeInternalStage) to ingest data into Snowflake using Snowpipe.

- SaaS processors

This release includes a large range of new SaaS processors allowing you to interact with systems such as Workday, Box, Google Drive, Dropbox, HubSpot, or Airtable.

- Registry client to connect to the Cloudera DataFlow Catalog

It is now possible to configure a Registry client in NiFi to connect to a Cloudera DataFlow Catalog instance and to use it as a NiFi Registry for checking out versioned flows and versioning your flows.

- Parameter Providers

It is now possible to create Parameter Providers in the Controller Settings. This gives the ability to source values of parameters from external sources such as HashiCorp Vault, AWS Secrets Manager, or Databases. You can read more about this new feature in *Using Parameter Providers*.

- Removed processors

Two processors have been removed in this release: ExecuteFlumeSink and ExecuteFlumeSource.

If you have a Flow Management Data Hub cluster and you want to upgrade from CDP 7.2.9 or lower, some manual steps are required before starting the Data Hub OS upgrade process. For more information, see [Performing a Data Hub OS upgrade](#).

What's New in Streams Messaging

Learn about the new Streams Messaging features in Cloudera DataFlow for Data Hub 7.2.16.

Kafka

Rebase on Kafka 3.1.2

Kafka shipped with this version of Cloudera Runtime is based on Apache Kafka 3.1.2. For more information, see the following upstream resources:

Apache Kafka Notable Changes:

- [3.0.0](#)
- [3.0.1](#)
- [3.1.0](#)
- [3.1.1](#)

Apache Kafka Release Notes:

- [3.0.0](#)
- [3.0.1](#)
- [3.1.0](#)
- [3.1.1](#)
- [3.1.2](#)

Multi-level rack awareness

The rack awareness capabilities of Kafka have been improved to support multi-level cluster topologies. As a result, brokers can now be configured to run in a multi-level rack-aware mode. If this mode is enabled, the brokers provide multi-level rack awareness guarantees. These guarantees ensure that topic partition replicas are spread evenly across all levels of the physical infrastructure. For example, in a two-level hierarchy with Data Centers on the top level and racks on the second level, brokers will evenly spread replicas among both available DCs and racks.

The new mode is compatible with follower fetching. If multi-level mode is enabled, a compatible replica selector class is automatically installed. This implementation enables consumers (if configured), to fetch Kafka messages from the replica that is closest to them in the multi-level hierarchy.

Additionally, when Cruise Control is deployed on the cluster, the standard rack-aware goals in Cruise Control's configuration are replaced with a multi-level rack-aware goal. This goal ensures that Cruise Control optimizations do not violate the multi-level rack awareness guarantees. This goal is currently downstream only, available exclusively in Cloudera distributed Cruise Control. For more information, see the following resources:

- [Kafka rack awareness](#)
- [Configure Kafka rack awareness](#)
- [Setting capacity estimations and goals](#)

Expose log directory total and usable space through the Kafka API

[KAFKA-13958](#) is backported in Kafka shipped with this version of Runtime. As a result, the Kafka API now exposes metrics regarding the total and usable disk space of log directories. The information on log directory space is collected by SMM and is exposed on the SMM UI. Specifically, you can now view the current log size of topics as well as the total log size and remaining storage space of brokers. For more information on how you can monitor log size metrics on the SMM UI, see [Monitoring log size information](#).

Kafka now accepts OAuth tokens that do not contain a “sub” claim

[KAFKA-13730](#) is backported in Kafka shipped with this version of Runtime. As a result, Kafka now accepts OAuth tokens that do not contain the "sub" claim. If you are using OAuth tokens that do not contain a “sub” claim, the JWT Principal Claim Name For OAuth2 Kafka service property must be configured. This property specifies the claim that contains the client's principal. For more information on OAuth2 authentication in Kafka, see [OAuth2 authentication](#).

New Kafka connect connectors

The following new Kafka connect connectors are introduced:

- HDFS Stateless Sink
- Influx DB Sink
- Debezium Db2 Source [Technical Preview]

For more information, see [Connectors](#).

Syslog TCP Source connector 2.0.0.

The Syslog TCP Source Kafka Connect connector is updated to version 2.0.0. The following notable changes and improvements are made:

- Three new properties are added, these are as follows:

- Max Batch Size

This property controls the maximum number of messages to add to a single batch of messages. This is a required property. Its default value is 1.

- Authorized Issuer DN Pattern and Authorized Subject DN Pattern

These properties allow you to enable authorization for incoming TLS connections. Both properties accept regular expressions as a value. The configured regular expressions are applied against the Distinguished Names of incoming TLS connections. If the Distinguished Names do not match the pattern, the following message is logged and the messages do not get forwarded to Kafka.

```
Error: authorization failure
```

Both properties are optional and are set to .* by default.

- The Max Number of TCP Connections property is replaced by the Max Number of Worker Threads property.

Similarly to Max Number of TCP Connections, Max Number of Worker Threads is also used to specify the number of TCP connections, but instead of exactly specifying the number of allowed connections, you now specify how many worker threads are reserved for TCP connections. Note that a single worker thread is capable of handling multiple connections. This is a required property. Its default value is 2.

- Existing version 1.0.0. connectors will continue to function, upgrading them, however, is not possible. If you want to use the new version of the connector, you must deploy a new instance of the connector.
- Deploying a version 1.0.0. instance of the connector is no longer possible.

AvroConverter support for KConnect logical types

The AvroConverter now converts between Connect and Avro temporal and decimal types.

Connect internal topic Ranger policy

A new Ranger policy, connect internal - topic, is generated by default on fresh installations. This policy allows the Kafka and SMM service principals to access Kafka Connect internal topics (connect-configs, connect-offsets, connect-status) and the secret management's storage topic (connect-secrets).

Connector configurations must by default override the sasl.jaas.config property of the Kafka clients used by the connector

The Require Connectors To Override Kafka Client JAAS Configuration Kafka Connect property is now selected by default. This means that connector configurations must by default contain a sasl.jaas.config entry with an appropriate JAAS configuration that can be used to establish a connection with the Kafka service.

Connect JAAS enforcement now applies to non-override type Kafka client configs

When the Require Connectors To Override Kafka Client JAAS Configuration property is selected, the consumer.sasl. and producer.sasl. configurations are not emitted into the Connect worker configurations anymore. Additionally, the keytab name is randomized and the \${cm-agent:keytab} references in the Connector configurations will stop working.

Schema Registry

Schema Registry instances behind load balancer

You can now use load balancer in front of Schema Registry instances. It is very common to have multiple instances of the same application and have a load balancer in front of them. This can be useful for failover reasons in HA environments, and it can also help sharing the load between

instances. You can also use load balancer in front of Schema Registry instances in an environment with Kerberos or SSL enabled.

AvroConverter support for KConnect logical types

AvroConverter now converts between Connect and Avro temporal and decimal types.

Support for alternative jersey connectors in SchemaRegistryClient

connector.provider.class can be configured in Schema Registry Client. If it is configured, schema.registry.client.retry.policy should also be configured to be different than default.

This also fixes the issue with some 3rd party load balancers where the client is expected to follow redirects and authenticate while doing that.

Retrieve principal from client's certificate

When two-way TLS authentication is enabled, Schema Registry extracts the principal from the certificate and uses it for authentication or authorization.

Schema Registry CDC support - change default schema compatibility

When a new Avro schema is created and its compatibility is not explicitly set, then a default compatibility value is used. Until now, that value was always BACKWARD. After this change, users on the server side can configure the default value.

Schema Registry with Knox uses round-robin load balancing

When multiple instances of Schema Registry are running, Knox uses round-robin to forward the requests.

Upgraded Avro version to 1.11.1

Avro got upgraded from version 1.9.1 to 1.11.1.

Streams Messaging Manager

Improved alter topic functionalities

You can now increase the number of partitions of a topic (but not decrease). The option is available on the Configs tab on the Topic Details page.

Partition Assignment tab on the Topic Details page

The Assignment tab, on the topic details page, shows the current state of the partitions and replicas of the topic. It shows some topic-level statistics and the replica assignment of all partitions. If rack awareness is being used in the Kafka cluster, the replica assignment is shown in a rack-based view. If the rack IDs follow the format of multi-level rack IDs, the rack IDs are rendered as a hierarchy.

Added sorter functionality to partition lists in SMM UI

The SMM UI contains Brokers and Topics pages where records contain broker or topic specific partition lists and their profile pages as well. All partition list columns become sortable.

SMM UI shows broker rack information

The Brokers page and the Broker Details page now both show the rack ID of the brokers. SMM now also supports a new endpoint: `/api/v1/admin/topics/{topicName}/description`.

Improved SMM UX for Kafka Connectors configuration

The connector selection and connector configuration workflow steps are now separated into two different steps. Search and autocomplete are now available for Connect configuration keys. The help icon provides detailed information about each configuration key. The data type of the configuration values can be chosen from the options menu. For more information, see *Setting connector configurations*.

Added partition log-size information to the SMM UI

The SMM UI shows log-size related information about brokers, topics, and partitions. Furthermore, warning messages appear when log directory related errors are reported by Kafka.

SMM connector profile shows connector level error

Errors causing the whole connector to fail are now displayed on the connector profile page if available.

Data explorer allows specifying consumer isolation.level

`/api/v1/admin/topics/{topicName}/partition/{partitionId}/payloads` endpoint has a new parameter: `consumerIsolationLevel`. The accepted values are `read_committed` and `read_uncommitted`. This sets the `isolation.level` config for the `KafkaConsumer` used for retrieving messages. The default value is `read_uncommitted`.

Additionally, the parameter can be set on the UI as well.

The Data Explorer page has a new design

The Data Explorer now uses the “from offset” and “record limit” parameters to select an offset window to query.

Offset-lag metrics in SMM UI

SMM UI Replications tab now also shows the replication-records-lag metric.

SMM UI Data Explorer shows null values explicitly

Data Explorer in the SMM UI now displays 'null' with italic style applied when the value is null rather than an empty value.

SMM uses a specific REST API to fetch list of topics

SMM Connect page now uses the Connect active topic tracking feature to list the topics used by the Connectors instead of the connector configuration's topic property. Sink Connectors show up based on which topics they consumed from (regardless of whether `topics` or `topics.regex` config was used), and Source Connectors show up based on which topics they produced to.

Improved configuration of SMM Kafka interceptors

New configuration prefix for SMM monitoring interceptor's background producer:
`smm.monitoring.interceptor.producer.`

Clients that use either of the SMM monitoring interceptors (`MonitoringConsumerInterceptor`, `MonitoringProducerInterceptor`) use a background producer to push client metrics into Kafka every 30 seconds. This background producer now can be configured by passing Producer Configurations to the client that uses the interceptor with the `smm.monitoring.interceptor.producer.` prefix. The prefix is trimmed and the remaining part of the configuration is passed to the background producer. For instance, if the user wishes to configure the `batch.size` property for the background producer, the following configuration should be passed: `smm.monitoring.interceptor.producer.batch.size`.

There is a different behavior, however, with the `client.id` property. If the user does not provide a configuration to the client id (`smm.monitoring.interceptor.producer.client.id`), the default is used, which is: `smm-monitoring-interceptor`.

SMM Data Explorer shows JSON output in pretty printed format

SMM Data Explorer can show JSON output in pretty printed format by using the `"JSON Pretty Print"` deserializer.

Streams Replication Manager

Internal SRM topics are now automatically added to the replication deny list

SRM now ignores internal SRM topics when replicating data. If required, this behavior can be disabled by adding `srm.internal.topic.exclude.enable=false` to Streams Replication Manager's Replication Configs in Cloudera Manager.

Metrics and health checks for the status processor Streams application in SRM Service

SRM Service health tests now show the state of the Connect status processor Streams application.

SRM topic creation timeout increased

Streams Replication Manager internal topic creation timeout property defaults are increased to 20s to tolerate intermittent issues at startup.

Replication status metrics are enhanced with task information

The SRM Service now reports the status of the connectors and tasks in the replication status with detailed descriptions. The status and detailed descriptions can be viewed on the Replications tab of SMM UI.

New metric replication-records-lag

The SRM Service now reports a new metric on its REST API called replication-records-lag. This metric provides information regarding the replication lag based on offsets. The metric is available both on the cluster and the topic level.

Raw metrics are compressed using LZ4

SRM internal metric producers from now on use LZ4 compression by default. LZ4 was chosen as it provides the best combination in terms of compression speed and performance. As a result, Cloudera recommends that you use LZ4. If required, however, you can change the compression by doing the following:

1. Add the following configuration entries to Streams Replication Manager's Replication Configs

```
workers.cloudera.metrics.reporter.producer.compression.type=[***COMPRESSION***]  
connectors.cloudera.metrics.reporter.producer.compression.type=[***COMPRESSION***]
```

2. Add the following to Additional Configs For Streams Application Running Inside SRM Service

```
producer.compression.type=[***COMPRESSION***]
```

For more information regarding, metrics, monitoring, as well as raw metric collection and aggregation, see [Streams Replication Manager monitoring and metrics](#).

Cruise Control

Rebasing Cruise Control to 2.5.85

Cruise Control is rebased to 2.5.85 version to be compatible with Apache Kafka 3.x version.

Adding MultiLevelRackAwareGoal to Cruise Control

As the currently available RackAwareGoal in Cruise Control does not support multi level rack awareness, a new goal was created to ensure that the replicas are assigned respecting the rules of multiple level racks.

The new goal is named `com.cloudera.kafka.cruisecontrol.analyzer.goals.MultiLevelRackAwareDistributionGoal`.

For more information, see the following resources:

- [Kafka rack awareness](#)
- [Configuring rack awareness](#)
- [Setting capacity estimations and goals](#)

What's New in Streaming Analytics

Learn about the new Streaming Analytics features in Cloudera DataFlow for Data Hub 7.2.16.



Warning: Upgrading to Streaming Analytics 7.2.16 is not supported. For information about how to migrate your Flink jobs to a cluster with a new version, see the [Migrating Flink jobs documentation](#).

7.2.16.1

The following new features are introduced in Streaming Analytics CDF for Data Hub 7.2.16.1:

Confluent Schema Registry support

Confluent Schema Registry is added as a supported Catalog in SQL Stream Builder.

Configurable column indexing for Materialized Views

When creating Materialized Views, you can enable column indexing and select the columns that need to be indexed.

7.2.16

The following new features are introduced in Streaming Analytics CDF for Data Hub 7.2.16:

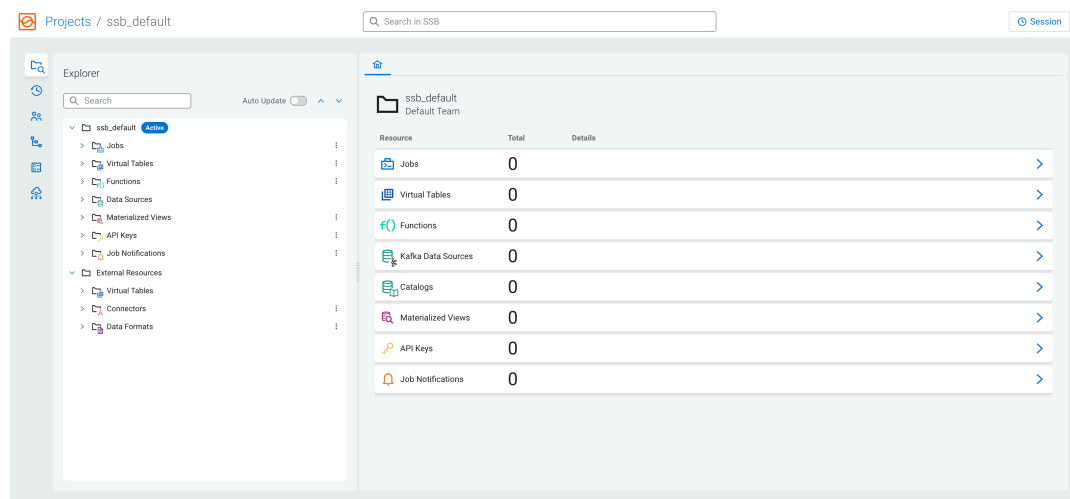
Apache Flink upgrade

Apache Flink 1.15.1 is supported in the Streaming Analytics 7.2.16 cluster definition.

For more information on what is included in the Apache Flink 1.15 version, see the [Apache Flink 1.15.1 Release Post](#) and the [Apache Flink 1.15 Release Notes](#)

Reworked Streaming SQL Console

The User Interface (UI) of SQL Stream Builder (SSB), the Streaming SQL Console has been reworked with new design elements.



Software Development Lifecycle (SDLC) support

Projects are introduced as an organizational element for SQL Stream Builder that allows you to create and collaborate on SQL jobs throughout the SDLC stages with source control. For more information, see the [Project structure and development](#) documentation.



Note: The Source Control and Service Discovery features are in Technical Preview and 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*.

Iceberg support

Support for Apache Iceberg table has been added as a connector to Flink. This allows you to create Iceberg tables in Streaming SQL Console.

For more information, see the [Creating Iceberg tables](#) documentation.



Note: The integration of Apache Iceberg with Flink and SQL Stream Builder is in Technical Preview and 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*.

Job notifications

You can enable notifications for SQL jobs to send information about job failures to a dedicated email group or through webhook.

For more information, see the [Using SQL job notification](#) documentation.

Diagnostic data collection

Data collection for diagnostic information is enabled for SQL Stream Builder to make troubleshooting and support processes easier.

For more information, see the [Collecting diagnostic data](#) documentation.

Materialized View pagination

You can set a limit and offset for the results of Materialized View queries to filter the query results more easier when accessing them through the Query API.

For more information, see the [Materialized View Pagination](#) documentation.

Configuring retention time for Materialized Views

You can configure how data should be retained for Materialized Views based on time or row count.

For more information, see the [Configuring Retention Time for Materialized Views](#) documentation.

SQL Server CDC Connector support

MS SQL Server Change Data Capture (CDC) connector is added to the list of supported connectors. The SQL Server CDC connector can be used with Flink SQL on the Streaming SQL Console either directly from the SQL Editor or using the DDL template.

JSON support for Schema Registry

JSON schemas are supported when using Cloudera Schema Registry with Kafka.

Component Support in Cloudera DataFlow for Data Hub 7.2.16

Cloudera DataFlow for Data Hub 7.2.16 includes the following components.

Flow Management clusters

- Apache NiFi 1.18.0.2.2.6.0
- Apache NiFi Registry 1.18.0.2.2.6.0



Note: Apache NiFi and Apache NiFi Registry version are unified in the 1.18.x release.

Streams Messaging clusters

- Apache Kafka 3.1.2
- Schema Registry 0.10.0
- Streams Messaging Manager 2.3.0
- Streams Replication Manager 1.1.0
- Cruise Control 2.5.85

Streaming Analytics clusters

- Apache Flink 1.15.1

Supported NiFi Extensions

Apache NiFi 1.18.0 ships with a set of Processors, Controller Services, Reporting Tasks, and Parameter Providers, 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 1.18.0 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 | GetHDFS | PutDistributedMapCache |
| AttributesToJSON | GetHDFSFileInfo | PutDynamoDB 1 |
| Base64EncodeContent | GetHDFSSequenceFile | PutDynamoDBRecord |
| CalculateRecordStats | GetHTMLElement | PutElasticsearchHttp 1 |
| CaptureChangeMySQL | GetHTTP | PutElasticsearchHttpRecord |
| CompressContent12 | GetHubSpot | PutElasticsearchJson |
| ConnectWebSocket | GetIgniteCache | PutElasticsearchRecord |
| ConsumeAMQP | GetJMSQueue | PutEmail 1 |
| ConsumeAzureEventHub | GetJMSTopic | PutFile |
| ConsumeEWS | GetMongoRecord | PutFTP |
| ConsumeGCPubSub | GetSFTP | PutGCSObject |
| ConsumeGCPubSubLite | GetShopify | PutGridFS |
| ConsumeJMS | GetSNMP | PutHBaseCell 1 |
| ConsumeKafka_1_0 | GetSnowflakeIngestStatus | PutHBaseJSON |
| ConsumeKafka_2_0 | (Technical Preview) | PutHBaseRecord |
| ConsumeKafka_2_6 | GetSolr | PutHDFS |
| ConsumeKafka2CDP | GetSplunk | PutHive3QL |
| ConsumeKafka2RecordCDP | GetSQS | PutHive3Streaming |
| ConsumeKafkaRecord_1_0 | GetTCP | PutHiveQL |
| ConsumeKafkaRecord_2_0 | GetTwitter | PutHiveStreaming |
| ConsumeKafkaRecord_2_6 | GetWorkdayReport | PutHTMLElement |
| ConsumeKinesisStream | GetZendesk | PutIceberg (Technical Preview) |
| ConsumeMQTT 1 | HandleHttpRequest | PutInfluxDB |
| ConsumeTwitter | HandleHttpResponse | PutJMS |
| ConsumeWindowsEventLog | HashAttribute | PutKinesisFirehose |
| ControlRate | HashContent | PutKinesisStream |
| ConvertAvroSchema | IdentifyMimeType | PutKudu |
| ConvertAvroToJSON | InvokeAWSGatewayApi | PutLambda |
| ConvertAvroToORC | InvokeGRPC | PutMongoRecord |
| ConvertAvroToParquet | InvokeHTTP | PutORC |
| ConvertCharacterSet | InvokeScriptedProcessor | PutParquet |
| ConvertCSVToAvro | JoinEnrichment | PutRecord |

| | | |
|----------------------------|---------------------------------|--|
| ConvertJSONToAvro | JoltTransformJSON | PutRiemann |
| ConvertJSONToSQL | JoltTransformRecord | PutS3Object |
| ConvertRecord | JsonQueryElasticsearch | PutSFTP |
| CreateHadoopSequenceFile | ListAzureBlobStorage | PutSmbFile |
| CryptographicHashAttribute | ListAzureBlobStorage_v12 | PutSnowflakeInternalStage (Technical Preview) |
| CryptographicHashContent | ListAzureDataLakeStorage | PutSNS |
| DecryptContentPGP | ListBoxFile | PutSolrContentStream |
| DeduplicateRecord | ListCDPObjectStore | PutSolrRecord |
| DeleteAzureBlobStorage | ListDatabaseTables | PutSplunk |
| DeleteAzureBlobStorage_v12 | ListDropbox | PutSplunkHTTP 1 |
| DeleteAzureDataLakeStorage | ListenBeats | PutSQL |
| DeleteByQueryElasticsearch | ListenFTP | PutSQS |
| DeleteCDPObjectStore | ListenGRPC | PutSyslog |
| DeleteDynamoDB | ListenHTTP | PutTCP |
| DeleteGCXObject | ListenRELP | PutUDP |
| DeleteGridFS | ListenSyslog | PutWebSocket 1 |
| DeleteHBaseCells | ListenTCP | QueryAirtableTable |
| DeleteHBaseRow | ListenTCPRecord | QueryCassandra |
| DeleteHDFS | ListenTrapSNMP | QueryDatabaseTable |
| DeleteS3Object | ListenUDP | QueryDatabaseTableRecord |
| DeleteSQS | ListenUDPRecord | QueryElasticsearchHttp |
| DetectDuplicate | ListenWebSocket | QueryRecord |
| DistributeLoad | ListFile | QuerySolr |
| DuplicateFlowFile | ListFTP | QuerySplunkIndexingStatus |
| EncryptContent 2 | ListGoogleDrive | QueryWhois |
| EncryptContentPGP | ListGCSBucket | ReplaceText 1 |
| EnforceOrder | ListGoogleDrive | ReplaceTextWithMapping |
| EvaluateJsonPath | ListHDFS | ResizeImage |
| EvaluateXPath | ListS3 | RetryFlowFile |
| EvaluateXQuery | ListSFTP | RouteHL7 |
| ExecuteGroovyScript | ListSmb | RouteOnAttribute |
| ExecuteInfluxDBQuery | LogAttribute | RouteOnContent |
| ExecuteProcess | LogMessage | RouteText |
| ExecuteScript | LookupAttribute | SampleRecord 1 |
| ExecuteSQL | LookupRecord | ScanAccumulo |
| ExecuteSQLRecord | MergeContent 1 | ScanAttribute |
| ExecuteStateless | MergeRecord | ScanContent |
| ExecuteStreamCommand | ModifyHTMLElement | ScanHBase |
| ExtractAvroMetadata | MonitorActivity | ScriptedFilterRecord |
| ExtractGrok | MoveAzureDataLakeStorage | ScriptedPartitionRecord |
| ExtractHL7Attributes | MoveHDFS | ScriptedTransformRecord |
| ExtractImageMetadata | Notify | ScriptedValidateRecord |
| ExtractText | PaginatedJsonQueryElasticsearch | ScrollElasticsearchHttp |
| FetchAzureBlobStorage | ParseCEF | SearchElasticsearch |
| FetchAzureBlobStorage_v12 | ParseEvtx | SegmentContent |
| FetchAzureDataLakeStorage | ParseSyslog | SelectHive3QL |
| FetchBoxFile | PartitionRecord | SelectHiveQL |
| FetchCDPObjectStore | PostHTTP | SendTrapSNMP |
| FetchDistributedMapCache | PublishAMQP 1 | SetSNMP |
| FetchDropbox | PublishGCPubSub 1 | |

| | | |
|------------------------|--------------------------|--|
| FetchElasticsearchHttp | PublishGCPubSubLite | SignContentPGP |
| FetchFile | PublishJMS 1 | SplitAvro 1 |
| FetchFTP | PublishKafka_1_0 | SplitContent 1 |
| FetchGCXObject | PublishKafka_2_0 | SplitJson 1 |
| FetchGoogleDrive | PublishKafka_2_6 | SplitRecord |
| FetchGridFS | PublishKafka2CDP | SplitText 1 |
| FetchHBaseRow | PublishKafka2RecordCDP | SplitXml 1 |
| FetchHDFS | PublishKafkaRecord_1_0 | StartSnowflakeIngest (Technical Preview) |
| FetchParquet | PublishKafkaRecord_2_0 | TagS3Object |
| FetchS3Object | PublishKafkaRecord_2_6 | TailFile |
| FetchSFTP | PublishMQTT 1 | TransformXml |
| FetchSmb | PutAccumuloRecord | UnpackContent |
| FlattenJson | PutAzureBlobStorage | UpdateAttribute |
| ForkEnrichment | PutAzureBlobStorage_v12 | UpdateByQueryElasticsearch |
| ForkRecord | PutAzureCosmosDBRecord 1 | UpdateCounter |
| GenerateFlowFile | PutAzureDataLakeStorage | UpdateDatabaseTable |
| GenerateTableFetch | PutAzureEventHub 1 | UpdateHive3Table |
| GeoEnrichIP | PutAzureQueueStorage | UpdateHiveTable |
| GeoEnrichIPRecord | PutBigQuery | UpdateRecord |
| GeohashRecord | PutBigQueryBatch | ValidateCsv |
| GetAzureEventHub | PutBigQueryStreaming 1 | ValidateRecord |
| GetAzureQueueStorage | PutCassandraQL 1 | ValidateXml |
| GetCouchbaseKey 1 | PutCassandraRecord | VerifyContentPGP |
| GetElasticsearch | PutCDPObjectStore | Wait |
| GetFile | PutCloudWatchMetric | YandexTranslate |
| GetFTP | PutCouchbaseKey 1 | |
| GetHBase | PutDatabaseRecord | |

Footnotes

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

Supported NiFi Controller Services

This release ships with Apache NiFi 1.18.0 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 | IPFIXReader |
| ActionHandlerLookup (Technical Preview) | IPLookupService |
| ADLSCredentialsControllerService | JASN1Reader |
| ADLSIDBrokerCloudCredentialsProviderControllerService | JMSConnectionFactoryProvider |
| AlertHandler (Technical Preview) | JndiJmsConnectionFactoryProvider |
| AvroReader | JsonConfigBasedBoxClientService |
| AvroRecordSetWriter | JsonPathReader |

| | |
|--|---|
| AvroSchemaRegistry | JsonRecordSetWriter |
| AWSCredentialsProviderControllerService | JsonTreeReader |
| AWSIDBrokerCloudCredentialsProviderControllerService | KafkaRecordSink_1_0 |
| AzureBlobIDBrokerCloudCredentialsProviderControllerService | KafkaRecordSink_2_0 |
| AzureCosmosDBClientService | KafkaRecordSink_2_6 |
| AzureEventHubRecordSink | KerberosKeytabUserService |
| AzureStorageCredentialsControllerService | KerberosPasswordUserService |
| AzureStorageCredentialsControllerService_v12 | KerberosTicketCacheUserService |
| AzureStorageCredentialsControllerServiceLookup | KeytabCredentialsService |
| CassandraDistributedMapCache | KuduLookupService |
| CassandraSessionProvider | LoggingRecordSink |
| CEFReader | LogHandler (Technical Preview) |
| CiscoEmblemSyslogMessageReader | MongoDBControllerService |
| ConfluentSchemaRegistry | MongoDBLookupService |
| CouchbaseClusterService | ParquetReader |
| CouchbaseKeyValueLookupService | ParquetRecordSetWriter |
| CouchbaseMapCacheClient | PrometheusRecordSink |
| CouchbaseRecordLookupService | ReaderLookup |
| CSVReader | RecordSetWriterLookup |
| CSVRecordLookupService | RecordSinkHandler (Technical Preview) |
| CSVRecordSetWriter | RecordSinkServiceLookup |
| DatabaseRecordLookupService | RedisConnectionPoolService |
| DatabaseRecordSink | RedisDistributedMapCacheClientService |
| DBCPCConnectionPool | RestLookupService |
| DBCPCConnectionPoolLookup | ScriptedActionHandler (Technical Preview) |
| DistributedMapCacheClientService | ScriptedLookupService |
| DistributedMapCacheLookupService | ScriptedReader |
| DistributedMapCacheServer | ScriptedRecordSetWriter |
| DistributedSetCacheClientService | ScriptedRecordSink |
| DistributedSetCacheServer | ScriptedRulesEngine (Technical Preview) |
| EasyRulesEngineProvider (Technical Preview) | SimpleDatabaseLookupService |
| EasyRulesEngineService (Technical Preview) | SimpleKeyValueLookupService |
| ElasticSearchClientServiceImpl | SimpleScriptedLookupService |
| ElasticSearchLookupService | SiteToSiteReportingRecordSink |
| ElasticSearchStringLookupService | SmbjClientProviderService |
| EmailRecordSink | SnowflakeComputingConnectionPool |
| EmbeddedHazelcastCacheManager | StandardDropboxCredentialService |
| ExpressionHandler (Technical Preview) | StandardHashiCorpVaultClientService |
| ExternalHazelcastCacheManager | StandardHttpContextMap |
| FreeFormTextRecordSetWriter | StandardOAuth2AccessTokenProvider |
| GCPCredentialsControllerService | StandardPGPPrivateKeyService |
| GrokReader | StandardPGPPublicKeyService |
| HadoopCatalogService (Technical Preview) | StandardPrivateKeyService |
| HadoopDBCPCConnectionPool | StandardProxyConfigurationService |
| HazelcastMapCacheClient | StandardRestrictedSSLContextService |
| HBase_1_1_2_ClientMapCacheService | StandardS3EncryptionService |
| HBase_1_1_2_ClientService | StandardSnowflakeIngestManagerProviderService |
| HBase_1_1_2_ListLookupService | (Technical Preview) |
| HBase_1_1_2_RecordLookupService | StandardSSLContextService |
| HBase_2_ClientMapCacheService | StandardWebClientServiceProvider |

| | |
|--|-----------------------|
| HBase_2_ClientService | Syslog5424Reader |
| HBase_2_RecordLookupService | SyslogReader |
| HiveCatalogService (Technical Preview) | UDPEventRecordSink |
| Hive3ConnectionPool | VolatileSchemaCache |
| HiveConnectionPool | WindowsEventLogReader |
| HortonworksSchemaRegistry | XMLReader |
| | XMLRecordSetWriter |

Supported NiFi Reporting Tasks

This release ships with Apache NiFi 1.18.0 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

Supported NiFi Parameter Providers

This release ships with Apache NiFi 1.18.0 and includes a set of Parameter Providers, most of which are supported by Cloudera Support. You should be familiar with the available supported Parameter Providers, and avoid using any unsupported Parameter Providers in production environments.

Additional Parameter Providers are developed and tested by the Cloudera community but are not officially supported by Cloudera. Parameter Providers 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.

- AwsSecretsManagerParameterProvider
- DatabaseParameterProvider
- EnvironmentVariableParameterProvider
- FileParameterProvider
- GcpSecretManagerParameterProvider
- HashiCorpVaultParameterProvider

Components supported by partners

This release ships with Apache NiFi 1.18.0 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 (v1.18.0)
- ConsumePulsarRecord (v1.18.0)
- PublishPulsar (v1.18.0)
- PublishPulsarRecord (v1.18.0)

Controller Services supported by partners

- PulsarClientAthenzAuthenticationService (v1.18.0)
- PulsarClientJwtAuthenticationService (v1.18.0)
- PulsarClientOAuthAuthenticationService (v1.18.0)
- PulsarClientTlsAuthenticationService (v1.18.0)
- StandardPulsarClientService (v1.18.0)

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 7.2.16

Some features exist within Cloudera DataFlow for Data Hub 7.2.16 components, but are not supported by Cloudera.

Unsupported Flow Management features

Some Flow Management features exist in Cloudera DataFlow for Data Hub 7.2.16, but are not supported by Cloudera.

NiFi

Technical preview features

The following features are available in Flow Management in Cloudera DataFlow for Data Hub 7.2.16 but are not ready for production deployment. Cloudera encourages you to explore these technical preview features in non-production environments and provide feedback on your experiences through the [Cloudera Community Forums](#).

- Controller Services:
 - ActionHandlerLookup
 - AlertHandler
 - EasyRulesEngineProvider
 - EasyRulesEngineService
 - ExpressionHandler
 - HadoopCatalogService
 - HiveCatalogService
 - LogHandler
 - RecordSinkHandler
 - ScriptedActionHandler
 - ScriptedRulesEngine
 - StandardSnowflakeIngestManagerProviderService
- Processors:
 - GetSnowflakeIngestStatus
 - PutIceberg
 - PutSnowflakeInternalStage
 - StartSnowflakeIngest

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 7.2.16, 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.
- 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](#)

[Setting up your Streams Messaging cluster](#)

Unsupported Streaming Analytics features

Some Streaming Analytic features exist in Cloudera DataFlow for Data Hub 7.2.16, but are not supported by Cloudera.



Warning: Upgrading to Streaming Analytics 7.2.16 is not supported. For information about how to migrate your Flink jobs to a cluster with a new version, see the [Migrating Flink jobs documentation](#).

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 7.2.16

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera DataFlow for Data Hub 7.2.16.

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:

Incomplete Ranger policy for NiFi metrics in Cloudera Manager

To have Cloudera Manager properly reflect the NiFi metrics for the NiFi service, the Flow NiFi access policies in Ranger needs to be updated to include the "nifi" group.

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

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
```

For more information, see [KUDU-3297](#).

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.

InferAvroSchema may fail when inferring schema for JSON data

In Apache NiFi 1.17, the dependency on Apache Avro has been upgraded to 1.11.0. However, the InferAvroSchema processor depends on the hadoop-libraries NAR from which the Avro version comes from, causing a NoSuchMethodError exception.



Important: This processor is not supported by Cloudera and its use is highly discouraged as inferring a schema from the data is not recommended in production data flows.

Having well defined schemas ensures consistent behavior, allows for proper schema versioning and prevents downstream systems to generate errors because of unexpected schema changes. Besides, schema inference may not always be 100% accurate and can be an expensive operation in terms of performances.

Use the ConvertRecord processor and have the Record Writer write the schema as a FlowFile attribute.

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

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-45183: Kafka Connect active topics might be visible to unauthorised users

The Kafka Connect active topics endpoint (`/connectors/[***CONNECTOR NAME***/topics)` and the Connect Cluster page on the SMM UI disregard the user permissions configured for the Kafka service in Ranger. As a result, all active topics of connectors might become visible to users who do not have permissions to view them. Note that user permission configured for Kafka Connect in Ranger are not affected by this issue and are correctly applied.

None.

OPSAPS-65485: Selecting the Require Connectors To Override Kafka Client JAAS Configuration property causes automatic Kafka Connect startup retries to fail

If the Require Connectors To Override Kafka Client JAAS Configuration property is selected for the Kafka Connect role and the role fails to start due to any reason, all automatic retries to start the role will also fail. This is true even if the root cause of the initial startup failure was caused by an intermittent issue.

Review the Kafka connect logs and resolve the root cause of the startup failure. Once the issue is resolved, restart the Kafka Connect role manually. Alternatively, clear the Require Connectors To Override Kafka Client JAAS Configuration property.

RANGER-3809: Idempotent Kafka producer fails to initialize due to an authorization failure

Kafka producers that have idempotence enabled require the Idempotent Write permission to be set on the cluster resource in Ranger. If permission is not given, the client fails to initialize and an error similar to the following is thrown:

```
org.apache.kafka.common.KafkaException: Cannot execute transactional method because we are in an error state
    at org.apache.kafka.clients.producer.internals.TransactionManager.maybeFailWithError(TransactionManager.java:1125)
    at org.apache.kafka.clients.producer.internals.TransactionManager.maybeAddPartition(TransactionManager.java:442)
```

```

    at org.apache.kafka.clients.producer.KafkaProducer.doSend(K
afkaProducer.java:1000)
    at org.apache.kafka.clients.producer.KafkaProducer.send(Kafk
aProducer.java:914)
    at org.apache.kafka.clients.producer.KafkaProducer.send(Kafk
aProducer.java:800)
    .
    .
    .
Caused by: org.apache.kafka.common.errors.ClusterAuthorization
Exception: Cluster authorization failed.

```

Idempotence is enabled by default for clients in Kafka 3.0.1, 3.1.1, and any version after 3.1.1. This means that any client updated to 3.0.1, 3.1.1, or any version after 3.1.1 is affected by this issue.

This issue has two workarounds, do either of the following:

- Explicitly disable idempotence for the producers. This can be done by setting `enable.idempotence` to `false`.
- Update your policies in Ranger and ensure that producers have Idempotent Write permission on the cluster resource.

CDPD-49304: AvroConverter does not support composite default values

AvroConverter cannot handle schemas containing a STRUCT type default value.

None.

DBZ-4990: The Debezium Db2 Source connector does not support schema evolution

The Debezium Db2 Source connector does not support the evolution (updates) of schemas. In addition, schema change events are not emitted to the schema change topic if there is a change in the schema of a table that is in capture mode. For more information, see [DBZ-4990](#).

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.

CDPD-48822: AvroConverter ignores default values when converting from Avro to Connect schema

AvroConverter does not propagate field default values when converting Avro schemas to Connect schemas.

None

Schema Registry

Learn about the known issues and limitations in Schema Registry in this release:

CDPD-48568: JAR storage does not work on AWS S3 for Schema Registry

If you are using AWS S3 to store the serializer and deserializer JAR files used by Schema Registry, the Schema Registry server might not be able reach these JAR files. As a result, message serialization and deserialization will not work with the affected schemas. If you are affected by this issue, class not found errors related to AWS library classes will be present in the Schema Registry server logs.

Ensure that the `aws-java-sdk-bundle-[***VERSION***].jar` file is available in `/opt/cloudera/parcels/CDH/lib/schemaregistry/hadoop-plugin/hadoop-schema-registry-plugin-impl/`. This is the location that contains all the JAR files for the Hadoop plugin classpath.

1. Find the `aws-java-sdk-bundle-[***VERSION***].jar` file.

The file is located in the folder that contains all common JAR files for parcels. You can find it using the following command:

```
find /opt/cloudera/parcels/CDH/jars -iname '*aws-java-sdk-bundle*.jar
```

The output of the command lists multiple versions. Take note of the latest version. For example:

```
/opt/cloudera/parcels/CDH/jars/aws-java-sdk-bundle-1.12.316.jar
```

2. Create a symlink in `/opt/cloudera/parcels/CDH/lib/schemaregistry/hadoop-plugin/hadoop-schema-registry-plugin-impl/` that points to the `aws-java-sdk-bundle-[***VERSION***].jar`. For example:

```
ln -s /opt/cloudera/parcels/CDH/jars/aws-java-sdk-bundle-1.12.316.jar /opt/cloudera/parcels/CDH/lib/schemaregistry/hadoop-plugin/hadoop-schema-registry-plugin-impl/aws-java-sdk-bundle-1.12.316.jar
```

3. Restart the Schema Registry service.

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-58265: Schema Registry Client incorrectly applies SSL configuration

The Cloudera distributed Schema Registry Java client might fail to apply the SSL configurations correctly with concurrent access in Jersey clients due to a [Jersey](#) issue related to JDK.

Before using `HttpsURLConnection` in any form concurrently, call `javax.net.ssl.HttpsURLConnection.getDefaultSSLSocketFactory()` once in the custom client application.

CDPD-48822: AvroConverter ignores default values when converting from Avro to Connect schema

`AvroConverter` does not propagate field default values when converting Avro schemas to Connect schemas.

None

CDPD-55381: Schema Registry issues authentication cookie for the authorized user, not for the authenticated one

When the authenticated user is different from the authorized user, which can happen when Schema Registry is used behind Knox, authorization issues can occur for subsequent requests as the authentication cookie in Schema Registry stores the authorized user.

Access Schema Registry directly, without using Knox, if possible. If not, ensure that the name of the end user that tries to connect does not begin with `knox`.

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-48853: Schemas created with the Confluent Schema Registry API cannot be viewed in the UI

Schemas created in Cloudera Schema Registry using the Confluent Schema Registry API are not visible in the Cloudera Schema Registry UI.

In addition, the `/api/v1/schemaregistry/search/schemas/aggregated` endpoint of the Cloudera Schema Registry API does not return schemas created with the Confluent Schema Registry API.

A typical case where this issue can manifest is when you are using the Confluent Avro converter for SerDes in a Kafka Connect connector and the connector connects to Cloudera Schema Registry. That is, the `key.converter` and/or `value.converter` properties of the connector are set to

`io.confluent.connect.avro.AvroConverter`, and `key.converter.schema.registry.url` and/or `value.converter.schema.registry.url` are set to a Cloudera Schema Registry server URL.

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 and limitations in Streams Messaging Manager in this release.

CDPD-39313: Some numbers are not rendered properly in SMM UI

Very large numbers can be imprecisely represented on the UI. For example, bytes larger than 8 petabytes would lose precision.

None.

CDPD-45183: Kafka Connect active topics might be visible to unauthorised users

The Kafka Connect active topics endpoint (`/connectors/[***CONNECTOR NAME***/topics`) and the Connect Cluster page on the SMM UI disregard the user permissions configured for the Kafka service in Ranger. As a result, all active topics of connectors might become visible to users who do not have permissions to view them. Note that user permission configured for Kafka Connect in Ranger are not affected by this issue and are correctly applied.

None.

CDPD-46465: Searching for workers on the connector overview page freezes the page

Navigating to the SMM UI Connectors page selecting the Cluster Profile tab and filtering for hosts using the search field causes the page to freeze.

None.

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-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.

Limitations

CDPD-36422: 1MB flow.snapshot freezes safari

Importing large connector configurations/ flow.snapshots reduces the usability of the Streams Messaging Manager's Connector page when using Safari browser.

Use a different browser (Chrome/Firefox/Edge).

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.

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

CDPD-60426: Configuration changes are lost following a rolling restart of the service

In certain cases, SRM might fail to apply configuration updates if the service is restarted with a rolling restart. In a case like this, configuration changes are ignored without any warning or indication. This issue also affects rolling upgrades.

When restarting the service, use **Actions Restart** instead of **Actions Rolling Restart** after making configuration changes. When upgrading a cluster, ensure that SRM is not restarted with a rolling restart.

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.

Cruise Control

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

Rebalancing with Cruise Control does not work due to the metric reporter failing to report the CPU usage metric

On the Kafka broker, the Cruise control metric reporter plugin may fail to report the CPU usage metric.

If the CPU usage metric is not reported, the `numValidWindows` in Cruise Control will be 0 and proposal generation as well as partition rebalancing will not work. If this issue is present, the following message will be included in the Kafka logs:

```
WARN com.linkedin.kafka.cruisecontrol.metricsreporter.CruiseControlMetricsReporter:
    [CruiseControlMetricsReporterRunner]: Failed reporting
    CPU util.
```

```
java.io.IOException: Java Virtual Machine recent CPU usage is not
    available.
```

This issue is only known to affect Kafka broker hosts that have the following specifications:

- CPU: Intel(R) Xeon(R) CPU E5-2699 v4 @ 2.20GHz
- OS: Linux 4.18.5-1.el7.elrepo.x86_64 #1 SMP Fri Aug 24 11:35:05 EDT 2018 x86_64
- Java version: 8-18

Move the broker to a different machine where the CPU is different. This can be done by performing a manual repair on the affected nodes. For more information, see the [Data Hub documentation](#).



Note: Cluster nodes affected by this issue are not displayed as unhealthy.

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.

- 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

7.2.16.1

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;
```

CSA-4541: Confluent Kafka cannot be validated

When using Confluent Kafka with SSB, the data source cannot be validated with SASL/SSL security method as the default truststore path is configured for CDP environments, and there is no option for custom parameters to be configured.

None

CSA-4426: Kafka Data Source name accepts spaces

Kafka Data Source can be validated and created with spaces in the Data Source name, but this results in errors as spaces are not valid characters based on the naming convention.

None

CSA-4464: CSA parcel is built with an interim CDP build

The CSA parcel is built using an interim CDP build and not with a build that corresponds to a release version. This can cause errors with components that have dependency to Flink.

If a Flink component transitively depends on some CDP related module and it is not accessible publicly, the dependency can be excluded. In case the project also depends on the excluded module, the publicly available version of the dependency can be added to Flink. For example, this could happen with `kafka-clients`, which is pulled in by `flink-connector-kafka`:

```
...
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-kafka</artifactId>
```

```

<version>${flink.version}</version>
<exclusions>
  <exclusion>
    <groupId>org.apache.kafka</groupId>
    <artifactId>kafka-clients</artifactId>
  </exclusion>
</exclusions>
</dependency>
<dependency>
  <groupId>org.apache.kafka</groupId>
  <artifactId>kafka-clients</artifactId>
  <version>3.1.2.7.2.16.0-287</version>
</dependency>
...

```

CSA-4426: Kafka Data Source name accepts spaces

Kafka Data Source can be validated and created with spaces in the Data Source name, but this results in errors as spaces are not valid characters based on the naming convention.

None

CSA-4412: Cannot delete Materialized View endpoint when using dynamic parameters

Materialized View endpoints cannot be deleted if dynamic parameters were set for them.

None

7.2.16.0**CSA-4650: Inconsistent sidebar collapse behavior**

The sidebar is collapsed inconsistently on the homepage of Streaming SQL Console when opening a project.

None

CSA-4643: flink-yarn-session is ignoring command line parameters

When adding parameters to the Flink session using `flink-yarn-session -d` in command line, the parameters are not applied to the session.

None

CSA-4548: Files cannot be uploaded through Swagger

REST API endpoints that take multipart requests, such as the uploading artifact endpoint, result in error.

None

CSA-4428: upsert-kafka template has an invalid property

The upsert Kafka template contains an invalid format property, which causes the table creation to fail.

None

CSA-4426: Kafka Data Source name accepts spaces

Kafka Data Source can be validated and created with spaces in the Data Source name, but this results in errors as spaces are not valid characters based on the naming convention.

None

CSA-4413: Missing AWS library

Due to the missing AWS library, SQL Stream Builder and Flink cannot be used in a RAZ-enabled environment.

None

CSA-4412: Cannot delete Materialized View endpoint when using dynamic parameters

Materialized View endpoints cannot be deleted if dynamic parameters were set for them.

None

CSA-4400: Cannot delete invalid catalog

A catalog can be created without a catalog service, but deleting the invalid catalog fails as it is not registered in Flink without a service.

None

CSA-4390: Missing flink-shaded-zookeeper-3 artifacts

The flink-shaded-zookeeper-3 artifacts are missing the public artifact repository.

Exclude the flink-shaded-zookeeper-3 dependency from the modules that would try to pull it in transitively. The following modules are relevant for the Flink application projects:

- flink-streaming-java
- flink-streaming-scala_2.12
- flink-table-api-java-bridge
- flink-table-api-scala-bridge_2.12
- flink-table-api-java-uber
- flink-clients, flink-statebackend-*
- flink-connector-*
- flink-sql-connector-*
- flink-test-utils
- flink-connector-test-utils
- flink-table-test-utils

CSA-4378: Validating Kafka Data Source fails without security settings

Kafka Data Source validation fails in a secure environment if the truststore details are not provided.

None

CSA-4377: Improve masking information of credentials

Credential information might be exposed on different configuration pages.

None

CSA-4372: Allow Deletion is not synchronized with project

When configuring a project with Allow Deletion setting and exported to a git repository, the Allow Deletion is not synchronized.

None

CSA-4370: Virtual tables imported from a schema in a Schema Catalog fail to describe correctly

Describing fails for a Virtual Table from a schema in a Schema Registry Catalog when viewing the table DDL.

None

CSA-4363: Production mode does not use clean session variables

When setting a variable in production mode, the job fails with an error as the variables set in production mode will not be used regardless of resetting the value of variable.

None

CSA-4360: Activating an environment does not reload catalogs

When activating an environment file in which variables are used for registered catalogs, the catalogs do not show up only after revalidation or after registering them again.

None

CSA-4356: Read-only jobs can be deleted

Running and read-only jobs can be deleted even though a member of a project does not have the permissions to stop and restart them.

None

CSA-4355: Virtual Table tab keeps loading for views

The Virtual Table tab is loading for an undefined time when opening a created view.

None

CSA-4354: Kudu lookup join fails

When using a Kudu lookup join and adding a WHERE clause with a string, the SSB job fails.

None

CSA-4353: Editing DDL Kafka tables opens wizard

When editing a Kafka table that was created using DDL, the configuration window appears as if the table was created using the Kafka Table wizard. This results in showing incorrect table information.

None

CSA-4352: Schema is not shown in table view

If a virtual table is opened in the table view, the schema is not shown and the information is displayed incorrectly.

None

CSA-4337: Materialized View link throws error

After importing a project and running the existing SQL job, the Materialized View query link does not work and the following error is shown:

```
"No data available yet. Verify that your Materialized View is configured, and the streaming job is running, and refresh page."
```

None

CSA-4333: Use Kafka Timestamps switch reflects invalid value

After creating a Kafka Virtual Table and disabling the "Use Kafka Timestamps" configuration, the table is created successfully according to the setting, but when viewing the DDL of the table, it shows the configuration as enabled.

None

CSA-4030: Webhook sending fails when webhook template is empty string

When creating a webhook table with a custom template, the webhook template will be saved as an empty string, which results in webhook sending failure.

None

CSA-3886: Session timeout configuration is not enforced

When setting the SSB Session Timeout configuration in Cloudera Manager, the set time is not enforced and SSB continues to run indefinitely.

None

CSA-3867: UDF case sensitivity

Due to a case sensitivity mismatch, User-Defined Functions (UDFs) can be created with uppercase, and be called by lowercase, but not vice-versa. This results in error when executing jobs as the UDFs are saved differently in the database.

None

CSA-3498: SSB fails with IllegalStateException

SSB fails to work and throws an IllegalStateException when running a job due to sample websocket session timeout.

None

Flink

7.2.16.0, 7.2.16.1

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;
```

In Cloudera Streaming Analytics, the following SQL API features are in preview:

- 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 7.2.16

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 7.2.16.

Fixed Issues in Flow Management

Review the list of Flow Management issues that are resolved in Cloudera DataFlow for Data Hub 7.2.16.

CFM 2.2.6.5-1

May 12 2023, CFM 2.2.6.5-1 hotfix

A hotfix has been released in the nifi-poi-nar-1.18.0.2.2.6.5-1.nar NAR file. This new NAR version resolves [NIFI-10792](#), so it is now possible to use ConvertExcelToCSVProcessor with files of over 10MB.

CFM 2.2.6

CFM 2.2.6 is rebased against Apache NiFi 1.18.0. It includes all fixed issues of this Apache NiFi release as well as the below list:

[NIFI-10660](#)

When converting Avro GenericRecord to a java Map, check GenericRecord's schema to see if field exists before calling Record.get().

[NIFI-10937](#) [NIFI-10938](#) [NIFI-10939](#)

When restoring flow from XML, properly parse XML for registry clients, looking for 'id' tag instead of 'identifier' tag. Fixed typo in error message about Reporting Task for Registry Clients. Updated StandardFlowComparator so that we compare sub-process groups' contents only if the Flow Coordinates are different by more than their version

[NIFI-10918](#)

When fetching a flow from a Flow Registry, if it references any 'internal versioned flows' instead of requiring that we have a client configured for the appropriate URL, attempt to fetch the flow from each client. We will start with the clients that do report that they can handle the URL but will try others as well. As soon as we successfully fetch the flow, we stop.

[NIFI-10914](#)

Adjusting input check for loading nested versioned flows

[NIFI-10901](#)

PublishKafka headers not sent in ProducerRecord

[NIFI-10874](#)

Fixed issue with multiple levels of versioned flows

[NIFI-10787](#)

Cannot commit flows to NiFi Registry after updating our NiFi release to 1.18.0. NifiRegistryFlowRegistryClient defines the PropertyDescriptor PROPERTY_URL name as small case "url". The map bases on the name property of the PropertyDescriptor object. Here searching with uppercase value of "URL" causes the map lookup to fail and cause a NPE later on. Therefore, it is changed as "url"

[NIFI-10817](#)

Moved the calls in StandardProcessSession to 'resetState()' and 'acknowledgeRecords()' from the outer commit(boolean) to the inner commit(Checkpoint, boolean). By moving the call here, the logic of StandardProcessSession is unaffected. But the StatelessProcessSession that inherits from it now has the benefit of having the state cleaned up when calling super.commit(Checkpoint, boolean).

NIFI-10814

Display description for selected flow

NIFI-10389

Added mime.type to GetShopify and GetHubSpot

NIFI-10828

Adding new REST API endpoint for exposing versioned flow details

NIFI-10793

Comment is not populated when process group is created via API

NIFI-9822

ConsumeKafkaRecord allows writing out Kafka record key

NIFI-10762

Ranger Authorizer accepts multiple Ranger Admin Identity property values

NIFI-10801

Fix hive metastore based tests in Iceberg processors

NIFI-10737

Corrected ListenBeats buffer handling

NIFI-4798

Allow empty value for UpdateAttribute property

NIFI-10130

AzureGraphUserGroupProvider handles group with transitive members

NIFI-10228

Parse json flow using instanceIdentifier as the UUID

NIFI-10287

ExecuteScript - Allow python scripts to use external modules

NIFI-10637

Simplified Locale validation in ParseCEF

NIFI-10675

Corrected SSL support in Neo4j V4 Cypher Service

NIFI-10764

Fixed handling BIT STRING in nifi-asn1-services

NIFI-10767

When an empty line is encountered with Syslog Readers, just skip over the empty line instead of throwing an Exception

NIFI-10768

Filter empty messages in ListenSyslog

NIFI-10771

Add SINGLE_BATCH_PER_NODE to allowable values for flow file concurrency in ProcessGroupDTO

NIFI-10783

Switched to LinkedHashMap for CompareFuzzyHash

NIFI-10788

Ensure proposed service config is applied when component synchronizer adds a new service

NIFI-10702

Clear server list on connection error in SMB processors

NIFI-10680

Corrected Bouncy Castle references in optional profiles

NIFI-10614

Upgraded Bouncy Castle from 1.70 to 1.71

NIFI-10698

Add filtering options to summary table

NIFI-10709

Showed log in link on logout complete page

NIFI-10317

- Taking care of NullPointerException if AMQP header value is null
- Refactoring to remove repeated conversions to string from char
- Correctly handle null values: null vs "null"
- Adding test
- Updated ConsumeAMQPTest to test for null and empty header values
- Updated ConsumeAMQPTest to use hard coded string values when testing
- Made convertMapToString() private

NIFI-9878

Added timeout handling for Cache Client handshaking

NIFI-10303

Route GetElasticsearch to failure if _id is blank after attribute evaluation

NIFI-10594

Render empty lines in Labels

NIFI-10631

Changed HashMap to LinkedHashMap in TestHBase_2_ClientService

NIFI-10635

Fix handling of enums in PutDatabaseRecord

NIFI-10641

Resolving typo on ParameterProvider warning message

NIFI-10643

Paging secrets in AwsSecretsManagerParameterProvider

NIFI-10668

Persist proxy password in remote process group

NIFI-10674

Block evaluateELString from reading Parameters

NIFI-10682

When determining effective property values, ensure that we always use the up-to-date version of Property Descriptors

NIFI-10687

Set Elasticsearch document `_id` to null if ID attribute evaluated to blank String for `PutElasticsearchRecord` or `PutElasticsearchJson`; use `@timestamp` default value if `@timestamp` record path evaluates to blank String in `PutElasticsearchRecord`

NIFI-10689

Fixed Parameter Provider Name encoding in context dialog

NIFI-10700

Only restart ports if their `scheduledState` is `Running`. - Adding unit tests. - Fixing unit test to properly verify Port isn't restarted.

NIFI-10693

Remove proxy configuration properties from `PutBigQuery`

NIFI-10673

When a component is added to a `ProcessGroup`, we had a method called `ensureUniqueVersionControlId` but the method only took into accounts in that Process Group. Updated methods to now consider components in any Process Group that is part of the same Versioned Flow. Also added system test to verify the problem and the fix. Added Toolkit clients and additional methods as necessary in order to implement system tests

NIFI-10644

Add Message Demarcator-style processing in `PublishMQTT`

NIFI-10628

Added Default Schedule Time and Run Duration to processors

NIFI-10650

Fix demarcator is appended to the end of the `FlowFile`'s content by `ConsumeMQTT`

NIFI-10567

Corrected Sensitive Dynamic Property handling for `flow.xml` (#6524)

NIFI-10648

Upgraded Apache Commons Text to 1.10.0

NIFI-10442

Create `PutIceberg` processor

NIFI-10602

Fix `GetHubSpot` and `GetShopify` state handling on NiFi restart

NIFI-10577

Fix Sensitive Checkbox rendering for `Fetch Parameters`

NIFI-10586

Prioritized `ssh-rsa` algorithm in `SFTP Processors`

NIFI-10597

Change 'App Config File' property to external resource in `Box` processors

NIFI-10273

Supported file entries larger than 8.5GB for `TAR` in `MergeContent`

NIFI-10543

Support broker failover in `MQTT` processors

CFM-2713

Resource Fetcher over `HTTPS` with Basic Authentication

NIFI-9955

Upgraded `ZooKeeper` 3.4.14 to 3.5.9 in Kite bundle

NIFI-9952

Adjusted Kite bundle to centrally managed jackson dependencies

NIFI-9933

Upgraded Apache Ant to 1.10.12

NIFI-10096

Corrected nested inherited Parameter Context loading

Fixed Issues in Streams Messaging

Review the list of Streams Messaging issues that are resolved in Cloudera DataFlow for Data Hub 7.2.16.

Kafka**CDPD-40985: Expose log directory total and usable space through the Kafka API (KAFKA-13958 backport)**

This is a backported improvement, see [KIP-827](#) and [KAFKA-13958](#) for more information.

CDPD-35680: Remove the verification of the existence of the JWT "sub" claim to extend compatibility with OAuth providers (KAFKA-13730 backport)

This is a backported improvement, see [KAFKA-13730](#) for more information.

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

The --zookeeper option has been removed from the kafka-topics tool. As a result, encountering this issue is no longer possible. Use the --bootstrap-server option instead.

Certain Kafka command line tools require direct access to ZooKeeper

There are no longer any Cloudera supported Kafka command line tools that require direct ZooKeeper access or require the usage of the --zookeeper option. Use the --bootstrap-server option instead.

CDPD-39422: Consumer polls for coordinator in tight loop (KAFKA-13917 backport)

The consumer uses a backoff period controlled by `retry.backoff.ms` to wait before polling for the coordinator if the previous poll was unsuccessful. This is a backported improvement, see [KAFKA-13917](#) for more information.

CDPD-39354: Kafka Connect connectors and tasks fail to start

Kafka Connect initializes the Secrets Storage right at startup. As a result, timing issues no longer occur in the configuration resolution of connectors and tasks.

CDPD-45958: Kafka client JAAS override policy validation is incorrect

The JAAS override filter policy now correctly filters based on the specified rules and does not refuse JAAS configurations because of unknown fields.

CDPD-44252: Exception during normal operation in MirrorSourceTask causes the task to fail instead of shutting down gracefully

Stopping the read of offsets in a worker of a MirrorSourceTask will now cause a graceful shutdown and the task can be restarted automatically at a later point.

CDPD-39391: Amazon S3 Sink fails when validating bucket names

Due to an issue with the AWS S3 bucket name validation of the Amazon S3 Sink connector, the connector encountered an exception when validating bucket names. This issue is now fixed.

OPSAPS-64606: Authorization issues if Kafka Connect is not installed

If the Kafka Connect role is not present on the cluster, then a Ranger policy (connect internal - topic) is created with default, non-empty topic names. As a result, the Ranger policy include list cannot be empty and will not have any side effects on other Kafka operations.

OPSAPS-63526: Kafka's Ranger related log4j2 config generation is not idempotent

If Kafka fails to start up due a configuration change, the original cause for the startup failure is no longer replaced with an error related to log4j2 configuration after the first auto-attempted restart.

OPSAPS-63640: Monitoring a high number of Kafka producers might cause Cloudera Manager to slow down and run out of memory

This issue is fixed.

Schema Registry**CDPD-20977: Add RAW Avro JSON Schema API for Hive Integration**

Added new endpoints where the client can GET the actual schema text as a JSON document. These endpoints were added as a subresource (.../schemaText) to the existing schema version resource endpoints, corresponding to the "schemaText" property of those, for example, /api/v1/schemaregistry/schemas/{name}/versions/latest/schemaText.

CDPD-39885: SchemaRegistryResource.uploadFiles fails w/ Timeout waiting for connection from pool

When downloading a file from Schema Registry, the stream was not properly closed which occasionally caused issues. This has been fixed by always closing the stream.

CDPD-40758: Improve upon how SchemaRegistryClient is used in connectors

Fixed the issue when setting the value.converter.serdes.protocol.version in connectors caused the connector to fail on startup because the configuration property was not properly converted to a byte value.

CDPD-41592: Confluent import should handle input without an actual version

Lines will be ignored in Confluent import format where only the first part corresponding to a Cloudera SR "meta" is present and schema version data is missing (or only a "null" string is there).

CDPD-45920: KafkaJsonDeserializer does not support primitive type literals and array types

Fixed JSON deserializer's handling of non-object literal values (simple strings, numeric literals or arrays).

Streams Messaging Manager**CDPD-33699: Remove "adjustTopicOverviewMetrics" from SMM**

Removed the logic introduced in 7.1.4/7.2.2.0, where in case the topicMetrics (bytes in/bytes out/messages in) are smaller for a larger time period, the smaller timeperiod's metrics will be displayed. For example, if the metrics are smaller for 30 days and then for 7 days, the 7 day metrics would be used.

CDPD-35136: Data Explorer should allow long keys to be viewed

Fixed the issue with the keys being truncated in the Data Explorer in case they are too long on SMM UI. In case the message's key is too long to render, a "read more" button will appear next to the truncated key, initiating a popup which will show the full content of the key.

CDPD-35374: Incorrect key/value for large numbers

Fixed issue where large numbers were rounded in the Data Explorer in case LongDeserialiser is used for either the key or the value while viewing the content of a partition.

CDPD-35897: Remove partitionMetrics from /api/v1/admin/metrics/consumers/group/{groupId} response

Removed the "topicPartitionMetrics" field from the response of "/api/v1/admin/metrics/consumers/group/{groupId}" endpoint, since the underlying metric has been removed. Additionally, the API path version was bumped to v2.

CDPD-36420: List icons should not appear for the validation errors

Removed bullet point from validation errors on the connector creation form of streams messaging manager UI.

CDPD-39137: Connect topic tracking line is not showing on UI

On the Connectors page, topic lineage lines are not showing up, only the line starts and ends. Now, Topic lineage lines are shown on the Connectors page.

CDPD-39778: Should not suggest flow.snapshot to be a secret

Streams Messaging Manager now does not allow the user to mark the flow.snapshot as sensitive data.

CDPD-39826: The Restart button for the ConnectorTasks is permanently disabled

Issue with the Restart button being permanently inactive on the ConnectorDetails page is fixed.

CDPD-39980: /api/v1/admin/auth/access throws NPE on unsecure environments

Fixed issue "/api/v1/admin/auth/access" throwing InternalServerError (Http code 500) when accessing the endpoint in a non-kerberized environment.

CDPD-40286: When using Oracle database, we can have an NPE during configuration

When using the Oracle database, the customer has the option to provide custom connection properties. Due to a bug, these connection properties were mandatory, which caused an error in case they were not provided.

CDPD-40758: Improve upon how SchemaRegistryClient is used in connectors

Fixed the issue when setting the "value.converter.serdes.protocol.version" in connectors caused the connector to fail on startup because the configuration property wasn't properly converted to a byte value.

CDPD-40871: SMM should fill "partitionMetrics" on "/api/v2/admin/metrics/aggregated/topics/{topicName}" even when partitionMetrics are empty

All the data that is available from the topic partitions are filled regardless whether metrics were fetched from Cloudera Manager or not.

CDPD-41069: Topic can be edited after selecting REPLICATION_STATUS as the alert attribute

Fixed the topic selection dropdown status in the alert editor after various UI events.

CDPD-41420: Schema version is not displayed in SMM when Avro value serializer is chosen

In the Data Explorer now it is possible to see all the schema versions associated with the given topic.

CDPD-41514: SMM displays CPU usage chart instead of CPU load

Fixed issue where SMM displays CPU usage chart instead of CPU load on the broker details page.

CDPD-41542: Consumer instance host field is empty

Fixed issue where consumer instance host field is empty in the SMM UI consumer details page.

CDPD-41553: Replication charts do not render on first visit

Fixed issue where replication charts do not render on first visit in the SMM topic details page.

CDPD-43387: Broker Details page does not show the Cloudera Manager button

The Cloudera Manager buttons that navigate to the broker resource within Cloudera Manager were not visible in previous releases. Now you can navigate from the SMM's broker view to the Cloudera Manager's broker view.

CDPD-43474: KConnect metrics cannot be scraped by prometheus in secure environments

Fixed Prometheus not being able to scrape Connect metrics from secure Kafka clusters. The "connect.prometheus.metrics.port" configuration was removed.

CDPD-43962: Performance improvement on the Broker Details page

The '/api/v2/admin/metrics/aggregated/brokers/{brokerId}' endpoint, called every time while opening the broker details page on the UI, was excessively slow when a large number of topics and partitions were present. This is now fixed by fetching partition metrics in bulk.

Streams Replication Manager

OPSAPS-63992: Rolling restart unavailable for SRM

Streams Replication Manager rolling restart support is added to Cloudera Manager. Streams Replication Manager can be restarted and upgraded without losing service availability.

Cruise Control

There are no fixed issues for Cruise Control in Cloudera DataFlow for Data Hub 7.2.16.

Fixed Issues in Streaming Analytics

Review the list of Streaming Analytics issues that are resolved in Cloudera DataFlow for Data Hub 7.2.16.

7.2.16.3

CSA-4541: Confluent Kafka cannot be validated

The issue regarding the validation error of Confluent Kafka is fixed.

CSA-4464: CSA parcel is built with an interim CDP build

The issue regarding CSA being built with an interim build is fixed.

FLINK-30966: Flink SQL IF FUNCTION logic error

The fix regarding the logic error in [FLINK-30966](#) is backported

7.2.16.1

CSA-4413: Missing AWS library

The issue regarding the missing AWS library is fixed, SSB and Flink can be used in a RAZ-enabled environment.

CSA-4390: Missing flink-shaded-zookeeper-3 artifacts

The missing flink-shaded-zookeeper-3 artifacts have been added to the public artifact repository.

CSA-4378: Validating Kafka Data Source fails without security settings

The issue regarding the Kafka Data Source validation without truststore information is fixed.

CSA-4377: Improve masking information of credentials

The issue regarding credential information being visible is fixed.

CSA-4372: Allow Deletion is not synchronized with project

The issue about not synchronizing Allow Deletion setting with exporting a project is fixed.

CSA-4363: Production mode does not use clean session variables

The issue about production mode not using clean session variables is fixed.

CSA-4360: Activating an environment does not reload catalogs

The issue regarding registered catalogs with activated environment not showing up is fixed.

CSA-4356: Read-only jobs can be deleted

The issue about deleting read-only jobs is fixed, jobs can only be deleted when stopped and by only those who have the permission to delete them.

CSA-4355: Virtual Table tab keeps loading for views

The issue about continuous loading when opening views in the Virtual Table tab is fixed.

CSA-4354: Kudu lookup join fails

The issue regarding SSB jobs failing due to Kudu lookup join with string in WHERE clause is fixed.

CSA-4353: Editing DDL Kafka tables opens wizard

The issue about editing Kafka tables that were created using DDL and opening the configuration in wizard view is fixed.

CSA-4352: Schema is not shown in table view

The issue about missing schemas when opening virtual tables in table view is fixed.

CSA-4337: Materialized View link throws error

The issue regarding the error when opening a Materialized View query link after importing is fixed.

CSA-3886: Session timeout configuration is not enforced

The issue about session timeout configuration not being enforced is fixed.

CSA-3867: UDF case sensitivity

The issue regarding case sensitivity for UDFs is fixed.

CSA-3498: SSB fails with IllegalStateException

The issue regarding SSB crashing with exception is fixed.

7.2.16

CSA-2016: Deleting table from other teams

The issue about deleting is not possible from other teams is fixed.

CSA-1231: Big numbers are incorrectly represented on the Streaming SQL Console UI

The issue regarding incorrectly represented numbers on Streaming SQL Console is fixed.

Fixed CVEs in Cloudera DataFlow for Data Hub 7.2.16

Review the list of CVEs that are resolved in Cloudera DataFlow for Data Hub 7.2.16.

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](#).

Fixed CVEs in Flow Management

Review the list of common vulnerabilities and exposures fixed in Cloudera Flow Management (CFM) in Data Hub 7.2.16.

No new fixed CVEs since CFM 2.2.5.0.



Note: CVE-2022-42889 - Apache Commons Text: Apache NiFi is not exposed to this vulnerability. However, in CFM 2.2.6, the dependencies on Apache Commons Text have all been upgraded to 1.10.0.

Behavioral Changes in Cloudera DataFlow for Data Hub 7.2.16

You can review the changes in certain features or functionalities of components that have resulted in a change in behavior from the previously released version to this version of Cloudera DataFlow for Data Hub 7.2.16.

Behavioral Changes in Streams Messaging

Review the list of Streams Messaging behavioral changes in Cloudera DataFlow for Data Hub 7.2.16.

Kafka

Summary:

The FileStream example connectors (FileStreamSourceConnector and FileStreamSinkConnector) are no longer available for deployment by default using the Kafka Connect Rest API, SMM UI, or SMM REST API. The JAR file of the connector is still shipped with Cloudera Runtime, but the connectors must be installed before they can be deployed. For more information on how to install Kafka Connect connectors, see [Installing Connectors](#).

Previous behavior:

The FileStream example connectors were available by default for deployment. In SMM, the connectors were selectable on the Connect ClusterConnector Setup page by default.

New behavior:

The FileStream example connectors must be installed before they can be deployed. In SMM, the connectors are no longer selectable on the Connect ClusterConnector Setup page by default.

Summary:

Topic auto creation for the internal consumers of the Kafka Connect workers is turned off by default.

Previous behavior:

Topic auto creation for the internal consumers of the Kafka Connect workers was turned on by default.

New behavior:

Topic auto creation for the internal consumers of the Kafka Connect workers is turned off by default.

Schema Registry

Summary:

Schema compatibility check now supports the evolution of Avro enums.

Previous behavior:

Schema compatibility check fails when a writer schema that has enum values is not present in the reader schema (some previous version), regardless if the reader has a default value.

New behavior:

A schema that does not have all enum values but has a default value set, remains compatible with a newer version that has an extended set of enum values.

Streams Messaging Manager

Summary:

The FileStream example connectors (FileStreamSourceConnector and FileStreamSinkConnector) are no longer available for deployment by default using the Kafka Connect Rest API, SMM UI, or SMM REST API. The JAR file of the connector is still shipped with Cloudera Runtime, but the connectors must be installed before they can be deployed. For more information on how to install Kafka Connect connectors, see [Installing Connectors](#).

Previous behavior:

The FileStream example connectors were available by default for deployment. In SMM, the connectors were selectable on the Connect ClusterConnector Setup page by default.

New behavior:

The FileStream example connectors must be installed before they can be deployed. In SMM, the connectors are no longer selectable on the Connect ClusterConnector Setup page by default.

Streams Replication Manager

Summary:

The default values for Metrics Topics Creation Timeout for Driver and Metrics Topics Creation Timeout for Service are increased to 20 seconds

Previous behavior:

The default values for Metrics Topics Creation Timeout for Driver and Metrics Topics Creation Timeout for Service was 5 seconds.

New behavior:

The default values for Metrics Topics Creation Timeout for Driver and Metrics Topics Creation Timeout for Service is 20 seconds.

Cruise Control

There are no behavioral changes for Cruise Control in Cloudera DataFlow for Data Hub 7.2.16.

Behavioral Changes in Streaming Analytics

Review the list of Streaming Analytics behavioral changes in Cloudera DataFlow for Data Hub 7.2.16.

7.2.16.1

Summary:

Loading columns in Materialized View **Query Editor** when initializing with active job

Previous behavior:

When creating a new Materialized View endpoint the **Query Editor** used the persisted SQL of the job to collect available columns even if the SQL statement was modified in the **SQL Editor**.

New behavior:

Materialized View **Query Editor** uses the current SQL query to collect available columns.

Summary:

Show dependencies when deleting Kafka Data Sources and UDFs

Previous behavior:

Deleting Kafka Data Sources and User-Defined Functions (UDFs) was possible without any warning message even if there were dependencies to other components.

New behavior:

Dependencies are shown and a confirmation message is displayed when deleting Kafka Data Sources and UDFs.

Summary:

Invalid catalog can be saved

Previous behavior:

Only validated catalogs could be saved in a project.

New behavior:

If a catalog is registered with a placeholder that is not activated yet with the environment, the catalog can be saved and validated later when the environment is activated.

7.2.16

Summary:

Some public API endpoints got reworked and have different endpoints or related models because of some business logic decisions.

Previous behavior:

For more information about the previous behavior, see the [previous version of the SQL Stream Builder API Reference documentation](#).

New behavior:

For more information about the new behavior, see the [latest version of the SQL Stream Builder API Reference documentation](#).