Cloudera DataFlow for Data Hub 7.2.15

# Cloudera DataFlow for Data Hub Release Notes

**Date published: 2019-12-16**
**Date modified: 2022-05-12**

**https://docs.cloudera.com/**

# Legal Notice

# Contents

# What's New in Cloudera DataFlow for Data Hub 7.2.15

Cloudera DataFlow for Data Hub 7.2.15 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 7.2.15.

Flow Management DataHub in CDP 7.2.15 is based on Apache NiFi 1.16.0 and includes significant improvements and fixes. Here are the most important new features and improvements:

- Retry at framework level

  This feature provides the ability to define a retry strategy for the flowfiles going into a given relationship for each processor.
- Clustering model and flow definition reconciliation

  A disconnected node in a NiFi cluster will no longer cause the UI to be read-only for flow changes and a reconciliation will be executed on the flow definition when the node rejoins the cluster.
- Snowflake Connection Pool controller service

  This controller service can be used in combination with the JDBC processors to easily interact with Snowflake for both pulling and pushing data.
- OAuth 2 integration with InvokeHTTP processor

  It is now possible to configure an OAuth2 token provider with InvokeHTTP to easily interact with services requiring OAuth2 authentication.
- The Kafka processors previously distributed for interacting with Kafka 0.x clusters have been removed. It means that these processors are no longer provided and supported by Cloudera. It is recommended to switch to the processors making use of Kafka 2.x versions.
  - ConsumeKafka
  - ConsumeKafka_0_10
  - ConsumeKafka_0_11
  - ConsumeKafkaRecord_0_10
  - ConsumeKafkaRecord_0_11
  - GetKafka
  - PublishKafka
  - PublishKafka_0_10
  - PublishKafka_0_11
  - PublishKafkaRecord_0_10
  - PublishKafkaRecord_0_11
  - PutKafka
- Elasticsearch processors that were leveraging the Elasticsearch 2.0 library and the deprecated Transport Client have been removed. It means that the following processors are no longer provided and supported by Cloudera. It is recommended to switch to the processors interacting with Elasticsearch over HTTP.
  - FetchElasticsearch
  - PutElasticsearch

## What's New in Streams Messaging

Learn about the new Streams Messaging features in Cloudera DataFlow for Data Hub 7.2.15.

## Streams Messaging cluster definitions and templates

### Streams Messaging High Availability cluster definition and template

Three new cluster definitions are introduced for Streams Messaging. The new definitions are as follows:

- Streams Messaging High Availability for AWS
- Streams Messaging High Availability for Azure (Technical Preview)
- Streams Messaging High Availability for Google Cloud (Technical Preview)

Additionally, a new template called CDP - Streams Messaging High Availability is also introduced. You can use the template and definitions to deploy highly available Streams Messaging clusters that leverage multiple availability zones and ensure that functionality is not degraded when a single availability zone has an outage. For more information regarding cluster layout, see Streams Messaging cluster layout. For more information on how to deploy a cluster with the new definition, see Creating your first Streams Messaging cluster.

## Kafka

### Enable JMX Authentication by default

JMX Authentication is now enabled by default for the Kafka service. Randomly generated passwords are now set for both the JMX monitor (read only access) and control (read and write access) users. The default passwords can be changed at any time using the Password of User with read-only Access to the JMX agent and the Password of user with read-write access to the JMX agent Kafka service properties. Additionally, JMX authentication can be turned off using the Enable Authenticated Communication with the JMX Agent property.

### OAuth2 authentication available for Kafka

Oauth2 authentication support is added for the Kafka service. You can now configure Kafka brokers to authenticate clients using Oauth2. For more information, see OAuth2 authentication.

### HSTS header is included by default in Kafka Connect REST API responses

Kafka Connect REST API responses now include the HSTS header by default.

### Kafka load balancer support

The Kafka service can now be provided with a host of a load balancer that is used to balance connection bootstraps between multiple brokers. The host can be configured using the Kafka Broker Load Balancer Host property. Additionally, if a host is configured, the Kafka service configures a listener for accepting requests from the load balancer. This port is customizable using the Kafka Broker Load Balancer Listener Port property. Using these properties configures your Kafka service in a way that clients can connect to the brokers without encountering ticket mismatch issues in Kerberized environments or TLS/SSL hostname verification failures.

### Importing Kafka entities into Atlas

Kafka topics and clients can now be imported into Atlas as entities (metadata) using a new action available for the Kafka service in Cloudera Manager. The new action is available at Kafka service>Actions>Import Kafka Topics Into Atlas. The action serves as a replacement/alternative for the kafka-import.sh tool. For more information, see Importing Kafka entities into Atlas.

### Debezium Connector support

The following change data capture (CDC) connectors are added to Kafka Connect:

- Debezium MySQL Source
- Debezium Postgres Source
- Debezium SQL Server Source
- Debezium Oracle Source

Each of the connectors require CDP specific steps before they can be deployed. For more information, see Connectors.

### Secure Kafka Connect

Kafka Connect is now generally available and can be used in production environments. This is the result of multiple changes, improvements, and new features related to Kafka Connect security including the following:

**SPNEGO authentication for the Kafka Connect REST API**

You can secure the Kafka Connect REST API by enabling SPNEGO authentication. If SPNEGO authentication is enabled, only users authenticated with Kerberos are able to access and use the REST API. Additionally, if Ranger authorization is enabled for the Kafka service, authenticated users will only be able perform the operations that they are authorized for. For more information, see Configuring SPNEGO Authentication and trusted proxies for the Kafka Connect REST API.

**Kafka Connect Authorization model**

An authorization model is introduced for Kafka Connect. Implementations are pluggable and it is up to the implementation how the capabilities of the model are utilized. The authorization model is implemented by default in Ranger. For more information about the model, see Kafka Connect authorization model. For more information about the Ranger integration of the model, see Kafka Connect Ranger integration.

**Kafka Connect connector configurations can now be secured**

A new feature called Kafka Connect Secrets Storage is introduced. This feature enables you to mark properties within connector configurations as a secret. If a property is marked as a secret, the feature stores and handles the value of that property in a secure manner. For more information, see Kafka Connect Secrets Storage.

**Kafka Connect Connectors can be configured to override the JAAS, and restrict the usage of the Worker principal**

Kafka Connect now allows users to force Connectors to override the JAAS configuration of the Kafka connection, and also forbid using the same Kerberos credentials as the Connect worker is using. For more information, see Configuring connector JAAS configuration and Kerberos principal overrides

**Nexus allow list for Stateless NiFi Source and Sink connectors**

A new configuration property, List Of Allowed Nexus Repository Urls, is introduced for the Kafka service. This property enables you to specify a list of allowed Nexus repositories that Kafka Connect connectors are allowed to connect to when fetching NiFi extensions. Configuring an allow list using the property can harden the security Kafka Connect deployment. For more information, see Configuring a Nexus repository allow list.

## Schema Registry

### Added OAuth support for Schema Registry client authentication

You can use OAuth2 JSON Web Token (JWT) in Schema Registry for authentication. Authorization continues to be implemented in Ranger, however, you can obtain the principal from a JWT token.

### Added a findAllSchemas() method to the Schema Registry Client code

Provides a findAllSchemas() method which enumerates all schemas contained in the schema registry, returned as a list of SchemaMetadataInfo. This is useful if you only need to enumerate all schemas by name, without incurring the additional overhead of the findAggregatedSchemas() method.

### Support for reading keys from JWK

Keys can be stored in JWK. The validation is done by matching with the "kid" property in JWT. If "kid" is not given then we match on the algorithm.

### Added JWT validation filter

Added Servlet filter which checks if the incoming requests contain a valid authentication token.

### SchemaRegistryClient gets token from OAuth Server with clientId/secret

Schema Registry Client can be configured to use OAuth2 authentication. The following parameters need to be added when creating a Schema Registry Client:

- "schema.registry.auth.type" = "oauth2" (default value is kerberos)
- "schema.registry.oauth.client.id" (ClientId for OAuth2 server)
- "schema.registry.oauth.secret" (Secret for OAuth2 server)
- "schema.registry.oauth.server.url" (REST API endpoint of OAuth2 server)

**Support added for RSA and HMAC certificates**

Added support for JWT signed by either RSA or HMAC.

## Streams Messaging Manager

**Improvement in the Connect tab of the SMM UI**

You can now deploy Kafka Connect connector configurations containing secret properties which will be stored in an encrypted storage (by default in Kafka). The deployed configuration will only contain references to these secrets. With this comes the need to mark properties as secret on the Streams Messaging Manager user interface so, a new connector creation form is introduced, which supports it. You can import configurations and populate the form automatically.

**Kafka Connect improvement**

- In NiFi connectors, you can now provide file path or URL for the flow.snapshot or alternatively you can upload it from file.
- You can now import Connector Configurations as a whole instead of adding individual configurations.
- Connector configuration validation errors are now correlated with individual config key.
- Sensitive properties are now hidden from the SMM UI and support is added to set properties as sensitive.

**Partition dimension removal in SMM**

The partition dimensions of the producer ("/api/v2/admin/metrics/aggregated/producers") and consumer ("/api/v2/admin/metrics/aggregated/groups") metrics are removed from the SMM cache, and are not exposed anymore through the API. This made the SMM memory footprint smaller, relieved some of the load from the metric store, and the network traffic became smaller. With this change, you get a cleaner, and easily readable API, and the UI is snappier, and faster than before.

The version of the /api/v1/admin/metrics/aggregated/* and /api/v1/admin/lineage/* endpoints have been changed to /api/v2/admin/aggregated and /api/v2/admin/lineage. With this change, the response objects are changed as well.

For the /lineage endpoints a common lineage response object is introduced in v2 as opposed to the specific (and different) objects in the experimental v1 endpoint.

For the /aggregated/* endpoints, the partition level metrics (that were in the wrappedPartitionMetrics field) are removed. Partition level metrics have been removed from the /aggregate/producers and /aggregated/producers/{producerClientId} but they are still available in the corresponding /metrics/producers and /metrics/producers/{producerId} endpoints.

**Stateless Sink and Source should populate Key/Value Converters**

SMM UI Connector Creation page now contains a default key/value converter to the StatelessNiFiSource or StatelessNiFiSink connectors.

**Added API to enrich a sample configuration**

Streams Messaging Manager API /connector-templates/config/enhance is added, which accepts a sample connector configuration and enhances it with the properties that are probably needed for that connector.

**Add "emit.consumer.metrics" config to SMM CSD, and remove (now) unused SMON host/port configs**

Removed "cm.metrics.service.monitor.host" and "cm.metrics.service.monitor.port" configurations from Streams Messaging Manager.

These no longer have to be configured as SMM automatically detects ServiceMonitor's location and emits the ConsumerGroup metrics into it.

Added "emit.consumer.metrics" configuration to Streams Messaging Manager.

In case this flag is disabled, Streams Messaging Manager does not emit historic ConsumerGroup metrics into ServiceMonitor, meaning historic metrics (for group Lag and CommittedOffset) would not be available for Groups in SMM. These metrics are used to populate the charts at the bottom of the ConsumerGroupDetail page, or accessed through the "api/v2/admin/metrics/consumers/group/{groupId}" REST API endpoint.

**Increase SMM version to 2.3**

SMM version is increased.

**SMM UI should show the replication status tooltip**

Streams Messaging Manager now shows tooltip for the replication status.

**On the Overview page adjust the lineage information shown**

On the Overview page, when a Producer or a Consumer is selected, an arrow points to the topic(s) it produced to or consumed from instead of the partitions.

### Streams Replication Manager

**SRM now creates all internal topics with correct configurations at startup**

The internal topics used by SRM are now automatically created with correct configurations at startup. These are the metrics topics, the topics used by the srm-control tool, and the topics used by the SRM Service for service discovery. Additionally, SRM also verifies that the topics are created with correct configurations. If the topics are not configured as expected, SRM fails to start. This improvement fixes CDPD-31745.

**Increase the default replication factor of internal topics to 3**

The internal topics used by SRM are now created with a replication factor of 3 by default. As a result, SRM is now more resistant to host failures. Additionally, Cruise Control can now automatically heal SRM’s internal topics in the event of a single host failure.

**SRM now waits for latest offset syncs and does not set the consumer offset into the future**

The MirrorCheckpointConnector now checks the latest message in the offset sync topic at startup, and does not emit a checkpoint message until it has read from the beginning all the messages prior and including that last message.

As a part of this improvement, a new configuration property, emit.checkpoints.end.offset.protection is introduced. When this property is enabled, the MirrorCheckpointTask checks the end offset of the replicated topic prior to emitting a checkpoint, and limits the replicated offset to be maximum that value. With this behavior enabled, SRM no longer encounters an issue where in certain situations the replicated offset could be higher than the end offset of the replicated topic, producing a negative lag. The property is enabled by default, but can be configured using the Streams Replication Manager's Replication Configs property.

### Cruise Control

**Configuration property for HTTP Strict Transport Security**

There is a new configuration property for Cruise Control that enables Strict Transport Security header in the web server responses when SSL is enabled. By default, the configuration is enabled, and when TLS is enabled, Cruise Control sets the Strict Transport Security policy in the web server responses.

## What's New in Streaming Analytics

Learn about the new Streaming Analytics features in Cloudera DataFlow for Data Hub 7.2.15.

The following new features are introduced in Streaming Analytics CDF for Data Hub 7.2.15:

### 7.2.15.2

**Configurable value for YARN queue**

The YARN queue can be configured for a job on the Streaming SQL Console using the SET statement, and the yarn.application.queue parameter.

For more information, see Configuring YARN application queue.

**Configuring data retention for Materialized Views**

You can configure how to retain data for a Materialized View based on time and data row.

For more information, see Configuring retention time for Materialized Views.

**Configuring checkpoints for SQL jobs**

You can configure checkpoints for SQL jobs to prevent data loss in case any error or failure occurs.

For more information, see Configuring SQL job settings.

### 7.2.15

**Configuration deserialization failures for Kafka tables**

Support for configuring error handling of deserialization is added. When using the Kafka connector with any data type, you can choose from the following options how to handle schema mismatch error:

- Throw an exception
- Ignore the message
- Ignore the message and log the error
- Ignore the message in the context of the current stream, but store it in a dead-letter queue topic

# Component Support in Cloudera DataFlow for Data Hub 7.2.15

Cloudera DataFlow for Data Hub 7.2.15 includes the following components.

Flow Management clusters

- Apache NiFi 1.16.0.2.2.5.0
- Apache NiFi Registry 1.16.0.2.2.5.0

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

Streams Messaging clusters

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

Streaming Analytics clusters

- Apache Flink 1.14

# Supported NiFi Extensions

Apache NiFi 1.16.0 ships with a set of Processors, Controller Services, and Reporting Tasks, most of which are supported by Cloudera Support. Review the supported extensions and avoid using any unsupported extensions in your production environments.

## Supported NiFi Processors

This release ships with Apache NiFi 1.16.0 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
AttributesToJSON
Base64EncodeContent
CalculateRecordStats
CaptureChangeMySQL
CompressContent12
ConnectWebSocket
ConsumeAMQP
ConsumeAzureEventHub
ConsumeEWS
ConsumeGCPubSub
ConsumeGCPubSubLite
ConsumeJMS
ConsumeKafka_1_0
ConsumeKafka_2_0
ConsumeKafka_2_6
ConsumeKafka2CDP
ConsumeKafka2RecordCDP
ConsumeKafkaRecord_1_0
ConsumeKafkaRecord_2_0
ConsumeKafkaRecord_2_6
ConsumeKinesisStream
ConsumeMQTT 1
ConsumeWindowsEventLog
ControlRate
ConvertAvroSchema
ConvertAvroToJSON
ConvertAvroToORC
ConvertAvroToParquet
ConvertCharacterSet
ConvertCSVToAvro
ConvertJSONToAvro
GetFTP
GetHBase
GetHDFS
GetHDFSFileInfo
GetHDFSSequenceFile
GetHTMLElement
GetHTTP
GetIgniteCache
GetJMSQueue
GetJMSTopic
GetMongoRecord
GetSFTP
GetSNMP
GetSolr
GetSplunk
GetSQS
GetTCP
GetTwitter
HandleHttpRequest
HandleHttpResponse
HashAttribute
HashContent
IdentifyMimeType
InvokeAWSGatewayApi
InvokeGRPC
InvokeHTTP
InvokeScriptedProcessor
JoinEnrichment
JoltTransformJSON
JoltTransformRecord
JsonQueryElasticsearch
ListAzureBlobStorage
PutDynamoDBRecord
PutElasticsearchHttp 1
PutElasticsearchHttpRecord
PutElasticsearchJson
PutElasticsearchRecord
PutEmail 1
PutFile
PutFTP
PutGCSObject
PutGridFS
PutHBaseCell 1
PutHBaseJSON
PutHBaseRecord
PutHDFS
PutHive3QL
PutHive3Streaming
PutHiveQL
PutHiveStreaming
PutHTMLElement
PutInfluxDB
PutJMS
PutKinesisFirehose
PutKinesisStream
PutKudu
PutLambda
PutMongoRecord
PutORC
PutParquet
PutRecord
PutRiemann
PutS3Object
PutSFTP

ConvertJSONToSQL
ConvertRecord
CreateHadoopSequenceFile
CryptographicHashAttribute
CryptographicHashContent
DecryptContentPGP
DeduplicateRecord
DeleteAzureBlobStorage
DeleteAzureBlobStorage_v12
DeleteAzureDataLakeStorage
DeleteByQueryElasticsearch
DeleteCDPObjectStore
DeleteDynamoDB
DeleteGCSObject
DeleteGridFS
DeleteHBaseCells
DeleteHBaseRow
DeleteHDFS
DeleteS3Object
DeleteSQS
DetectDuplicate
DistributeLoad
DuplicateFlowFile
EncryptContent 2
EncryptContentPGP
EnforceOrder
EvaluateJsonPath
EvaluateXPath
EvaluateXQuery
ExecuteGroovyScript
ExecuteInfluxDBQuery
ExecuteProcess
ExecuteScript
ExecuteSQL
ExecuteSQLRecord
ExecuteStateless
ExecuteStreamCommand
ExtractAvroMetadata
ExtractGrok
ExtractHL7Attributes
ExtractImageMetadata
ExtractText
FetchAzureBlobStorage
FetchAzureBlobStorage_v12
FetchAzureDataLakeStorage
FetchCDPObjectStore
FetchDistributedMapCache
FetchElasticsearchHttp
FetchFile
FetchFTP
ListAzureBlobStorage_v12
ListAzureDataLakeStorage
ListCDPObjectStore
ListDatabaseTables
ListenFTP
ListenGRPC
ListenHTTP
ListenRELP
ListenSyslog
ListenTCP
ListenTCPRecord
ListenTrapSNMP
ListenUDP
ListenUDPRecord
ListenWebSocket
ListFile
ListFTP
ListGCSBucket
ListHDFS
ListS3
ListSFTP
LogAttribute
LogMessage
LookupAttribute
LookupRecord
MergeContent 1
MergeRecord
ModifyHTMLElement
MonitorActivity
MoveAzureDataLakeStorage
Notify
PaginatedJsonQueryElasticsearch
ParseCEF
ParseEvtx
ParseSyslog
PartitionRecord
PostHTTP
PublishAMQP 1
PublishGCPubSub 1
PublishGCPubSubLite
PublishJMS 1
PublishKafka_1_0
PublishKafka_2_0
PublishKafka_2_6
PublishKafka2CDP
PublishKafka2RecordCDP
PublishKafkaRecord_1_0
PublishKafkaRecord_2_0
PublishKafkaRecord_2_6
PublishMQTT 1
PutSNS
PutSolrContentStream
PutSolrRecord
PutSplunk
PutSplunkHTTP 1
PutSQL
PutSQS
PutSyslog
PutTCP
PutUDP
PutWebSocket 1
QueryCassandra
QueryDatabaseTable
QueryDatabaseTableRecord
QueryElasticsearchHttp
QueryRecord
QuerySolr
QuerySplunkIndexingStatus
QueryWhois
ReplaceText 1
ReplaceTextWithMapping
ResizeImage
RetryFlowFile
RouteHL7
RouteOnAttribute
RouteOnContent
RouteText
SampleRecord 1
ScanAccumulo
ScanAttribute
ScanContent
ScanHBase
ScriptedFilterRecord
ScriptedPartitionRecord
ScriptedTransformRecord
ScriptedValidateRecord
ScrollElasticsearchHttp
SearchElasticsearch
SegmentContent
SelectHive3QL
SelectHiveQL
SendTrapSNMP
SetSNMP
SignContentPGP
SplitAvro 1
SplitContent 1
SplitJson 1
SplitRecord
SplitText 1
SplitXml 1

FetchGCSObject
FetchGridFS
FetchHBaseRow
FetchHDFS
FetchParquet
FetchS3Object
FetchSFTP
FlattenJson
ForkEnrichment
ForkRecord
GenerateFlowFile
GenerateTableFetch
GeoEnrichIP
GeoEnrichIPRecord
GeohashRecord
GetAzureEventHub
GetAzureQueueStorage
GetCouchbaseKey 1
GetElasticsearch
GetFile
PutAccumuloRecord
PutAzureBlobStorage
PutAzureBlobStorage_v12
PutAzureCosmosDBRecord 1
PutAzureDataLakeStorage
PutAzureEventHub 1
PutAzureQueueStorage
PutBigQueryBatch
PutBigQueryStreaming 1
PutCassandraQL 1
PutCassandraRecord
PutCDPObjectStore
PutCloudWatchMetric
PutCouchbaseKey 1
PutDatabaseRecord
PutDistributedMapCache
PutDynamoDB 1
TagS3Object
TailFile
TransformXml
UnpackContent
UpdateAttribute
UpdateByQueryElasticsearch
UpdateCounter
UpdateHive3Table
UpdateHiveTable
UpdateRecord
ValidateCsv
ValidateRecord
ValidateXml
VerifyContentPGP
Wait
YandexTranslate

Footnotes

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

## Supported NiFi Controller Services

This release ships with Apache NiFi 1.16.0 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
ActionHandlerLookup
ADLSCredentialsControllerService
ADLSIDBrokerCloudCredentialsProviderControllerService
AlertHandler
AvroReader
AvroRecordSetWriter
AvroSchemaRegistry
AWSCredentialsProviderControllerService
AWSIDBrokerCloudCredentialsProviderControllerService
AzureBlobIDBrokerCloudCredentialsProviderControllerService
AzureCosmosDBClientService
AzureStorageCredentialsControllerService
AzureStorageCredentialsControllerService_v12
HortonworksSchemaRegistry
IPFIXReader
IPLookupService
JASN1Reader
JMSConnectionFactoryProvider
JndiJmsConnectionFactoryProvider
JsonPathReader
JsonRecordSetWriter
JsonTreeReader
KafkaRecordSink_1_0
KafkaRecordSink_2_0
KafkaRecordSink_2_6
KerberosKeytabUserService
KerberosPasswordUserService

AzureStorageCredentialsControllerServiceLookup
CassandraDistributedMapCache
CassandraSessionProvider
CEFReader
CouchbaseClusterService
CouchbaseKeyValueLookupService
CouchbaseMapCacheClient
CouchbaseRecordLookupService
CSVReader
CSVRecordLookupService
CSVRecordSetWriter
DatabaseRecordLookupService
DatabaseRecordSink
DBCPConnectionPool
DBCPConnectionPoolLookup
DistributedMapCacheClientService
DistributedMapCacheLookupService
DistributedMapCacheServer
DistributedSetCacheClientService
DistributedSetCacheServer
EasyRulesEngineProvider
EasyRulesEngineService
ElasticSearchClientServiceImpl
ElasticSearchLookupService
ElasticSearchStringLookupService
EmailRecordSink
EmbeddedHazelcastCacheManager
ExpressionHandler
ExternalHazelcastCacheManager
FreeFormTextRecordSetWriter
GCPCredentialsControllerService
GrokReader
HadoopDBCPConnectionPool
HazelcastMapCacheClient
HBase_1_1_2_ClientMapCacheService
HBase_1_1_2_ClientService
HBase_1_1_2_ListLookupService
HBase_1_1_2_RecordLookupService
HBase_2_ClientMapCacheService
HBase_2_ClientService
HBase_2_RecordLookupService
Hive3ConnectionPool
HiveConnectionPool
KerberosTicketCacheUserService
KeytabCredentialsService
KuduLookupService
LoggingRecordSink
LogHandler
MongoDBControllerService
MongoDBLookupService
ParquetReader
ParquetRecordSetWriter
PrometheusRecordSink
ReaderLookup
RecordSetWriterLookup
RecordSinkHandler
RecordSinkServiceLookup
RedisConnectionPoolService
RedisDistributedMapCacheClientService
RestLookupService
ScriptedActionHandler
ScriptedLookupService
ScriptedReader
ScriptedRecordSetWriter
ScriptedRecordSink
ScriptedRulesEngine
SimpleDatabaseLookupService
SimpleKeyValueLookupService
SimpleScriptedLookupService
SiteToSiteReportingRecordSink
SnowflakeComputingConnectionPool
StandardHttpContextMap
StandardOauth2AccessTokenProvider
StandardPGPPrivateKeyService
StandardPGPPublicKeyService
StandardProxyConfigurationService
StandardRestrictedSSLContextService
StandardS3EncryptionService
StandardSSLContextService
Syslog5424Reader
SyslogReader
VolatileSchemaCache
WindowsEventLogReader
XMLReader
XMLRecordSetWriter

## Supported NiFi Reporting Tasks

This release ships with Apache NiFi 1.16.0 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

## Components Supported by Partners

This release ships with Apache NiFi 1.16.0 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 (v1.15.2)
- ConsumePulsarRecord (v1.15.2)
- PublishPulsar (v1.15.2)
- PublishPulsarRecord (v1.15.2)

**Controller Services supported by partners**

- PulsarClientAthenzAuthenticationService (v1.15.2)
- PulsarClientJwtAuthenticationService (v1.15.2)
- PulsarClientOauthAuthenticationService (v1.15.2)
- PulsarClientTlsAuthenticationService (v1.15.2)
- StandardPulsarClientService (v1.15.2)

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 7.2.15

Some features exist within Cloudera DataFlow for Data Hub 7.2.15 components, but are not supported by Cloudera.

# Unsupported Flow Management features

There are no unsupported Flow Management features in Cloudera DataFlow for Data Hub 7.2.15

## NiFi

There are no updates for this release.

## NiFi Registry

There are no updates for this release.

### Related Information

Cloudera Community Forum

# Unsupported Streams Messaging features

Some Streams Messaging features exist in Cloudera DataFlow for Data Hub 7.2.15, 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

Creating your first Streams Messaging cluster

## Unsupported Streaming Analytics features

Some Streaming Analytic features exist in Cloudera DataFlow for Data Hub 7.2.15, but are not supported by Cloudera.

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

### Flink

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

### Related Information

Cloudera Community Forum

# Known Issues In Cloudera DataFlow for Data Hub 7.2.15

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera DataFlow for Data Hub 7.2.15.

## Known Issues in Flow Management

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

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

**KafkaRecordSink puts multiple records in one message**

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

For more information, see NIFI-8326.

There is no workaround for this issue.

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

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

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

There is no workaround for this issue.

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

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

Update the configuration of the ReportLineageToAtlas reporting task:

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

### Parameter Context inheritance may be lost during NiFi restart

Upon restarting NiFi, the inheritance between parameter contexts may be lost under specific conditions. It is recommended to upgrade to the latest version or to request a HOTFIX via the support portal.

For more information, see NIFI-10096.

## Technical Service Bulletins

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

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

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

#### Upstream JIRA

- NIFI-10023
- NIFI-9993

#### Knowledge article

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

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

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

#### Knowledge article

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

# Known Issues in Streams Messaging

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

## Kafka

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

Known Issues

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

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

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

**Certain Kafka command line tools require direct access to Zookeeper**

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

- kafka-reassign-partitions

None

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

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

None

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

The first few produce requests fail when sending to a nonexistent topic with auto.create.topics.e nable 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-63640: Monitoring a high number of Kafka producers might cause Cloudera Manager to slow down and run out of memory**

This issue has two workarounds. You can either configure a Kafka producer metric allow list or completely disable producer metrics.

- Configure a Kafka producer metric allow list:

  A producer metric allow list can be configured by adding the following properties to Kafka Broker Advanced Configuration Snippet (Safety Valve) for kafka.properties.

  ```
  producer.metrics.whitelist.enabled=true
  producer.metrics.whitelist=[***ALLOW LIST REGEX***]
  ```

  Replace *[***ALLOW LIST REGEX***]* with a regular expression matching the client.id of the producers that you want to add to the allow list. This regular expression uses the java.util.regex. Pattern class to compile the regular expression, and uses the match() method on the client.id to determine whether it fits the regular expression.

  Once configured, the metrics of producers whose client.id does not match the regular expression provided in producer.metrics.whitelist are filtered.Kafka no longer reports these metrics through the HTTP metrics endpoint. Additionally, existing metrics of the producers whose client.id does not match the regular expression are deleted.

  Because the allow list filters metrics based on the client.id of the producers, you must ensure that the client.id property is specified in each producer's configuration. Automatically generated client IDs might cause the number of unnecessary metrics to increase even if an allow list is configured.
- Completely disable producer metrics:

  Producer metrics can be completely disabled by unchecking the Enable Producer Metrics Kafka service property.

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

AvroConverter cannot handle schemas containing a STRUCT type default value.

None.

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

The JAAS override filter policy refuses configurations if the configuration contains an unknown field instead of only refusing based on known fields with invalid values.

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 HostsHosts 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

**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 err
or occurred while processing Atlas events.
java.lang.IllegalArgumentException: Cannot invoke com.hortonworks
.registries.schemaregistry.AtlasEventStorable.setType on bean cl
ass 'class com.hortonworks.registries.schemaregistry.AtlasEventS
torable' - argument type mismatch - had objects of type "java.la
ng.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-39826: The Restart button for the ConnectorTasks is permanently disabled**

On the ConnectorDetails page, the Restart button for the tasks within the connector is permanently disabled.

Restart the whole Connector.

**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-separat
ed 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.

**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-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.

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 [TARGE
T_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, it's 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.

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

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

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

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

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

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

## Cruise Control

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

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

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

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

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

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

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

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

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

## Technical Service Bulletins

**TSB 2022-614: Kafka policy "user auto-creation" does not work in Ranger in CDP Public Cloud 7.2.15**

Creating a completely new Cloudera Data Platform (CDP) Public Cloud 7.2.15 Streams Messaging Light Duty Data Hub cluster fails after the Data Lake upgrade to 7.2.15. Note that the Data Hub cluster is created, but it looks unusable because of the lack of permissions.

New user principals are added to Apache Kafka (Kafka) policies (cc_metric_reporter and kafka_mirror_maker) in CDP Public Cloud version 7.2.14 as part of new features. Whenever a new Data Hub cluster is installed, its Kafka service is started for the first time, it will try to create all the default Kafka related policies automatically. If any of the users referred to in the policies does not exist in Apache Ranger (Ranger), it will refuse creating any of the policies in CDP Public Cloud version 7.2.15 for the new Data Hub cluster and the cluster will look unusable.

This is because from CDP Public Cloud 7.2.15 onwards, Ranger only lets administrators create new users. Automatic user creation works in CDP Public Cloud 7.2.14, so the affected customers will depend on which versions of Streams Messaging Data Hub clusters and Data Lakes they used earlier and how they used them.

**Knowledge article**

For the latest update on this issue see the corresponding Knowledge article: TSB 2022-614: Kafka policy "user auto-creation" does not work in Ranger in CDP Public Cloud 7.2.15

## Known Issues in Streaming Analytics

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

### SQL Stream Builder

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

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

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

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

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

**Uploading connector files fail**

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

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

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

**Upgrading Streaming Analytics cluster to 7.2.15**

Due to missing information in the database of SQL Stream Builder, upgrading the Streaming Analytics clusters to 7.2.15 is not possible.

None

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

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

None

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

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

Use DROP TABLE statement from the SQL window.

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

You must consider the timezone settings of your environment when using timestamps in a Kafka table as it can affect the results of your query. When the timestamp in a query is identified with

from_unixtime, it returns the results based on the timezone of the system. If the timezone is not set in UTC+0, the timestamp of the query results will shift in time and will not be correct.

Change your local timezone settings to UTC+0.

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

The issue impacts the following scenarios in Streaming SQL Console:

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

None

## Flink

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

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

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

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

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

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 7.2.15

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 7.2.15.

## Fixed Issues in Flow Management

Review the list of Flow Management issues that are resolved in Cloudera DataFlow for Data Hub 7.2.15.

### 7.2.15.2

#### Technical Service Bulletins

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

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

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

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

### 7.2.15

**NIFI-9943**

Added Transform Provider to nifi-xml-processing.

**NIFI-9901**

- Removed provided scope from nifi-xml-processing registry-ranger.
- Added nifi-xml-processing to nifi-commons.

**NIFI-9835**

Fixed threading bug in which NioAsyncLoadBalanceClient calls LoadBalanceSession.isComplete() followed by LoadBalanceSession.isCanceled() but it's possible for the complete flag to change before the canceled flag (they are not updated atomically). So changed to use a single LoadBalanceSessionState enum that represents the state. Also made the private StandardProcessSession.commit(boolean) method synchronized. When a processor is terminated (as is the case in Offload), we roll back sessions and both the commit() and rollback() need to be synchronized. Only the public commit() method was synchronized, and now with commitAsync() happening, we had the ability to commit without any synchronization. This addresses that concern. Also fixed a typo in docs for MergeRecord.

**NIFI-9834**

When calling ByteArrayContentRepository.read() on a null Content Claim, return an empty ByteArrayInputStream instead of throwing NullPointerException.

**NIFI-9827**

Upgrading AWS Java SDK to 1.12.182 to pick up new AWS Regions.

**NIFI-9818**

Fix flaky tests.

**NIFI-9815**

Corrected log message formatting in multiple classes.

**NIFI-9807**

Added Refresh Window Property to OAuth2 Token Provider.

**NIFI-9806**

Introduce ConfigurableExtensionDefinition and VersionedConfigurableExtension (#5875).

**NIFI-9801**

Fixed error in previous correction of AccessToken.isExpired() margin calculation NIFI-9801 Stabilized shaky AccessTokenTest.

**NIFI-9800**

Unwrap SQLException in PutDatabaseRecord when table does not exist.

**NIFI-9799**

Enabled style checking for nifi-system-tests in ci-workflow.

**NIFI-9797**

Corrected AccessToken.isExpired() margin calculation.

**NIFI-9796**

Updated Registry Security Configuration to avoid warnings.

**NIFI-9795**

Checkstyle, rat issues in nifi-system-test-suite module.

**NIFI-9794**

If a node is OFFLOADING, do not allow connections to be deleted. Also if we fail to notify the node that it needs to offload its data, change its state back to DISCONNECTED.

**NIFI-9791**

Use maven.build.timestamp during manifest generation instead of buildhelper timestamp.

**NIFI-9790**

Fixed race condition in which SwappablePriorityQueue could attempt to access the 0th element of an empty collection.

**NIFI-9789**

Upgraded Logback from 1.2.10 to 1.2.11.

**NIFI-9788**

Updated commons-codec to 1.15 across all modules.

**NIFI-9786**

Added debug to KeyStoreUtils.isStoreValid.

**NIFI-9785**

Improved Login Credentials Writer File Handling.

**NIFI-9783**

When migrating FlowFiles from one ProcessSession to another, if any FlowFile had already been transferred, and the Relationship to which it was transferred was auto-terminated, we were updating the wrong member variable, which threw off our stats for the processor. Fixed that.

**NIFI-9782**

Excluded H2 DB from nifi-druid-bundle.

**NIFI-9781**

Fix handling when selecting array element via QueryRecord.

**NIFI-9778**

Fixing additional details for ScriptedPartitionRecord.

**NIFI-9777**

Adding support to remove attributes from verification requests.

**NIFI-9775**

Create RuntimeManifestService.

**NIFI-9774**

Upgraded Netty from 4.1.73 to 4.1.74.

**NIFI-9771**

When a Kafka record is obtained during config verification, we should produce an invalid response if the Record Reader is not able to produce any records from it.

**NIFI-9766**

Avoid intermittent SearchElasticsearchTest failures in CI pipeline.

**NIFI-9765**

Added documentation that covers how to build a custom binar…

**NIFI-9764**

Atlas reporting task sends 'unknown' hive_table when table is name not available.

**NIFI-9763**

- Additional escaped VALUE column due to H2 changes.
- Escaped VALUE column for Configure Details Auditing.

**NIFI-9762**

Adding DBCPConnectionPool config verification.

**NIFI-9761**

Correct PeerChannel processing for TLS 1.3.

**NIFI-9759**

Upgraded Spring Framework from 5.3.15 to 5.3.16.

**NIFI-9757**

Upgraded SLF4J from 1.7.35 to 1.7.36.

**NIFI-9756**

Add documentation for framework-level retry in Processors and update processor tab images.

**NIFI-9754**

Introduced VersionedExternalFlow - Updated stateless and StandardProcessGroup, etc. to make use of VersionedExternalFlow - Updated StatelessDataflowDefinition to use ExternalVersionedFlow instead of generic type - Updated Stateless Bootstrap to avoid loading stateless engine libs from root classpath but instead use a NarClassLoader to load the stateless nar.

**NIFI-9751**

Poll as needed during system-tests to ensure expected state.

**NIFI-9750**

Logging Improvements to support LoadBalanceProtocol troubleshooting.

**NIFI-9749**

Capture additional logging for system-test workflow runs.

**NIFI-9748**

Added new property for Output Format to LogAttribute. Also made the FlowFile Properties (file size, entry date, lineage start date) optional and renamed from 'Standard FlowFile Attributes' to 'FlowFile Properties' because this has led to confusion many times in the past, around users wanting to reference these things as attributes via EL but they are not actually attributes.

**NIFI-9747**

Track PID in nifi-bootstrap logging on shutdown.

**NIFI-9745**

Prevent insertion of revisions in NiFi registry when revision feature is disabled.

**NIFI-9743**

Upgraded Jetty from 9.4.44 to 9.4.45.

**NIFI-9741**

Make the close() method of WriteAvroResultWithExternalSchema idempotent.

**NIFI-9738**

VersionedComponent data members should derive from Object.

**NIFI-9736**

Improved TestRouteText to avoid intermittent failures.

**NIFI-9735**

Corrected Jetty Duplicate Mapping Warning.

**NIFI-9734**

Standardized exception cause message formatting.

**NIFI-9733**

Fixing StandardConnection.verifySourceStoppedOrFunnel infinite recursion.

**NIFI-9732**

Upgraded Zip4j from 2.8.0 to 2.9.1.

**NIFI-9731**

Updated to use a shorter, simpler output format for FlowFiles when creating bulletins.

**NIFI-9730**

Consider a change in value for retry-related fields from 'null' to the default value as an environmental change so that it's not flagged as a Local Modification, which would prevent users from updating the version of the Process Group that they are using.

**NIFI-9729**

When restarting components in the VersionedFlowSynchronizer, first filter out any components that are intended to be stopped.

**NIFI-9728**

Added support for User Assigned Managed Identity authentication for Azure ADLS and Blob_v12 processors.

**NIFI-9727**

IndexOutOfBoundsException in CorrelationAttributePartitioner.

**NIFI-9726**

Removed duplicate nifi-utils dependencies from graph modules.

**NIFI-9725**

On shutdown, instead of spawning a background thread to shutdown Cluster Coordinator, do so in the calling thread. This avoids a race condition whereby the cluster coordinator cannot be determined because the other thread has shutdown the FlowController.

**NIFI-9724**

Added set-sensitive-properties-algorithm command.

**NIFI-9723**

When we add controller-level Controller Services on restart of NiFi, ensure that all Controller Services are updated to include their property values, etc. Also ensure that for these services and reporting tasks we decrypt the property values.

**NIFI-9722**

Do not throw an Exception from verifyCanUpdateProperties when property descriptor & parameter descriptor's sensitivities don't match - instead allow the set to happen and let processor become invalid. Also, allow values such as abc#{param} for ghost processors.

**NIFI-9721**

Support enum types in AvroTypeUtil.buildAvroSchema().

**NIFI-9716 and NIFI-9577**

Addressed issue in the PathFilter for GetFile / ListFile. For any file that is found in the Input Directory directly, it was previously being listed/fetched even if it didn't match the PathFilter. Additionally, updated the code to create a new File FIlter for every invocation of onTrigger. This was necessary for NIFI-9577 because the directory to monitor supports Expression Language and as a result may change from invocation to invocation, if using a function such as now() but the PathFilter would always relativize the path based on the value that was obtained when the processor was scheduled.

**NIFI-9715**

Add option to output empty FlowFile from Elasticsearch REST API Json Query processors when there are no hits from query.

**NIFI-9714**

Added overloaded toMap to MapRecord that can convert sub-records into maps.

**NIFI-9713**

TagS3Object do not have provenance data.

**NIFI-9711**

Added support for flow.json.gz in SetSensitivePropertiesKey.

**NIFI-9707**

Resolved duplicate JLine dependency in nifi-toolkit.

**NIFI-9704**

Updated the ContentRepositoryScanTask to show details of how much content in the content repo is retained by each queue in the dataflow. Changed default for nifi.content.claim.max.appendable.size property from 1 MB to 50 KB. Updated docs to reflect the new default value and explain what the property does and how it's used.

**NIFI-9699**

Updated oidcCallback method to handle error cases. Added some unit tests.

**NIFI-9698**

When creating an Avro schema, ensure that any default value is converted from what is returned by RecordField.getDefaultValue() to what Avro requires.

**NIFI-9696**

DeleteS3Object don't have provenance.

**NIFI-9692**

Upgraded Apache Commons Lang3 to 3.12.0.

**NIFI-9691**

Added ForkEnrichment, JoinEnrichment processors.

**NIFI-9689**

- When checking FlowFile Availability, consider swap queue and trigger data to be swapped in, since calling poll() will no longer happen if no data is available.
- When all FlowFiles in a FlowFile Queue are penalized, do not schedule the destination to run. Also expose this fact via the ConnectionStatusSnapshotDTO, as this allows the front-end to render this information to the user in order to avoid confusion when it appears that the Processor has data but does nothing.

**NIFI-9688**

Improve Logback shutdown handling.

**NIFI-9687**

Add additional documentation for nifi.cluster.node.protocol.max.threads property to Admin Guide.

**NIFI-9686**

Renamed SNMP integration tests correctly.

**NIFI-9685**

Upgraded JNA to 5.10.0.

**NIFI-9684**

Fix: When starting/stopping a selected process group, it sends the parent process group id to the REST interface that is responsible to enable/disable transmission for all remote process groups within a process group. Need to send the id of the select process group instead.

**NIFI-9681**

Upgraded Apache Commons DBCP to 2.9.0.

**NIFI-9679**

Added access-environment-credentials permission.

**NIFI-9678**

Update Elasticsearch REST API processor integration-tests for Elasticsearch 8.x.

**NIFI-9673**

Improved DBCP and HikariCP test reliability.

**NIFI-9672**

Fix flaky tests caused by the use of HashMap.

**NIFI-9669**

Adding PutDynamoDBRecord processor.

**NIFI-9668**

Adding informative error when setting same value in 'set-param' CLI command.

**NIFI-9663**

Setting the "csv-escape" property has no effect in SelectHive3QL.

**NIFI-9662**

Remove unused mail-1.4.7 dependency from nifi-framework-bundle.

**NIFI-9660**

Upgraded Apache Tika to 2.3.0.

**NIFI-9657**

Create MoveADLS processor.

**NIFI-9655**

Add Queue Logging to ListenUDP.

**NIFI-9652**

Upgraded jetty-schemas from 3.1 to 5.2.

**NIFI-9650**

Upgraded OkHttp from 4.9.2 to 4.9.3.

**NIFI-9649**

Upgraded SLF4J from 1.7.32 to 1.7.35.

**NIFI-9647**

Added ExtractDocumentText Processor.

**NIFI-9645**

Updated PutSplunk to allow idle connection timeouts.

**NIFI-9644**

Improved TestExecuteStateless increasing TestRunner.run() wait allowed.

**NIFI-9642**

Update Admin Guide and User Guide with correct nifi.properties default values.

**NIFI-9641**

Adjusted the extraction of the chroot suffix for solr client connections.

**NIFI-9639**

Determine how long it takes to find cluster coordinator and perform DNS lookup when sending heartbeats and include in the logs.

**NIFI-9638**

- Refactored Google Guava usage in extensions.
- Refactored Google Guava references.

**NIFI-9635**

Upgraded Netty from 4.1.72 to 4.1.73

**NIFI-9634**

Upgraded Spring Framework to 5.3.15.

**NIFI-9632**

Removed nifi-lumberjack-bundle.

**NIFI-9631**

Enable cli.sh to be used with a symbolic link.

**NIFI-9630**

Migrated Registry REST API docs to swagger-codegen.

**NIFI-9629**

Ensure that when we are setting default values on Avro GenericRecord objects that we convert from the schema's default value to the proper type.

**NIFI-9628**

Added a uiOnly flag when requesting Controller Service.

**NIFI-9626**

Allowing Stateless NiFi to parse flow snapshots with unrecognized fields.

**NIFI-9625**

- Refactored Distributed Cache Server and Client Tests - Replaced TestServerAndClient with separate classes for Set Server and Map Server - Implemented before and after annotations for starting and stopping server instances.
- Added check for cache directory existence before clean NIFI-9625 Updated Map and Set Cache Server Tests to use random port.

**NIFI-9624**

Removed JCenter Repository.

**NIFI-9621**

Added Ignore Reserved Characters to FlattenJson.

**NIFI-9620**

Adding isStateful to StatelessDataflow.

**NIFI-9619**

Removed GPG key from Security Mailing List reporting.

**NIFI-9618**

Upgraded Checkstyle to 9.2.1.

**NIFI-9617**

Removed unused screenshots from documentation.

**NIFI-9616**

Included SLF4J bridge libraries in NiFi Stateless Kafka Connect assembly.

**NIFI-9611**

- Restore commons-io to minifi-assembly.
- Removed unnecessary references to nifi-processor-utils.
- Removed duplicate nifi-utils dependency.

**NIFI-9610**

Refactored nifi-processor-utils to separate modules.

**NIFI-9609**

Added nifi-snowflake-bundle with SnowflakeComputingConnectionPool.

**NIFI-9608**

Disabled system-tests workflow for pull requests.

**NIFI-9607**

Honor Update Keys when Quoting Identifiers in PutDatabaseRecord.

**NIFI-9606**

Removed nifi-security-utils from nifi-framework-api.

**NIFI-9601**

Upgraded nifi-bootstrap to Jakarta Mail 2.

**NIFI-9600**

Removed Elasticsearch 2 Processors.

**NIFI-9599**

Updating explicit plugin/build version references.

**NIFI-9597**

- Fixing all explicit version refs to main latest.

- Fix Dockerfile URLs.

**NIFI-9596**

Fix newline bug in JythonScriptRunner NIFI-9596: Added comment to indicate why the Apache header is missing.

**NIFI-9595**

Removed nifi-kafka-0.x modules.

**NIFI-9594**

When converting Record to Avro GenericRecord, ensure that any default values that are defined in the GenericRecord's schema get applied, regardless of whether or not the field exists in the associated RecordSchema.

**NIFI-9593**

Missing catch clauses in Confluent Schema Registry client.

**NIFI-9591**

Removed nifi-kite-bundle.

**NIFI-9590**

Added support for sensitive properties in Azure authorizers to encrypt-config.

**NIFI-9589**

Support initial loading from the current max values in QueryDatabaseTable* processors.

**NIFI-9588**

Update doc for 'nifi.content.repository.archive.max.retention.period'.

**NIFI-9587**

Added JSON format for Prometheus Flow Metrics.

**NIFI-9586**

- Removed Surefire ForkNodeFactory configuration.
- Excluded additional assembly modules from ci-workflow.

**NIFI-9585**

Upgraded H2 from 1.4 to 2.1.210.

**NIFI-9581**

Add PutElasticsearchRecord relationship for output of successful Records sent to Elasticsearch.

**NIFI-9580**

UI work for framework-level retry in Processors.

**NIFI-9576**

- Introduced a BlockListClassLoader that can be used by stateless in order to isolate both the Stateless Engine and the NiFi extensions from extraneous classes that exist in the System ClassLoader.
- Allowed Stateless' BlockListClassLoader to load java11 jars/classes.

**NIFI-9575**

Updating copyright year to 2022.

**NIFI-9571**

Corrected Session commit handling in PutTCP.

**NIFI-9570**

Separate classpath for NiFi Registry sensitive property providers.

**NIFI-9569**

SNMP manager UDP transportmapping changed to 0.0.0.0.

**NIFI-9568**

Updated nifi-jolt-transform-json-ui pom.xml to only include CSS and JS assets in WAR.

**NIFI-9564**

Removed unnecessary logback-classic test dependencies.

**NIFI-9563**

Enabled ListenTCP Pool Receive Buffers property.

**NIFI-9552**

Fix NoClassDefFound error in case of nifi-registry-ranger-assembly.

**NIFI-9548**

When disabling RPG transmission, wait for the ports to complete in a background thread instead of blocking the web thread. Also moved the RPG initialization logic into flow controller instead of flow service and added a delay in order to reduce likelihood of ConnectException happening when pointing to nodes in the same cluster.

**NIFI-9545**

Fix in-place replacement for LookupRecord processor.

**NIFI-9544**

LookupRecord - fixed behavior when no matching value in the LRS.

**NIFI-9543**

Add bring-to-front functionality to labels.

**NIFI-9527**

Upgraded snappy-java to 1.1.8.4.

**NIFI-9525**

Modify lib packaging to use files from build directory.

**NIFI-9501**

Added REST end-point to retrieve a RuntimeManifest.

**NIFI-9481**

Excluded Data Transfer REST methods from DoSFilter.

**NIFI-9475**

Provide Framework-Level Retries for NiFi Relationships.

**NIFI-9455**

Added aggregated predictions to Prometheus Flow Metrics.

**NIFI-9453**

Refactored ListenBeats using Netty.

**NIFI-9438**

Refactored sensitive-property-provider to multiple modules.

**NIFI-9435**

Added registries and names include parameters to Flow Metrics.

**NIFI-9425**

Added auto-load NAR capability to MiNiFi.

**NIFI-9400**

Ensure that we always use the CollectionUsage metrics in the mbeans instead of the Usage metrics.

**NIFI-9390**

Updates to MergeContent / MergeRecord so that they play nicely within Stateless.

**NIFI-9348 and NIFI-7863**

Added temporary suffix and fixed NIFI-7863 creation of the directories.

**NIFI-9341**

- Corrected annotation syntax problem.
- Added CEF RecordReader.

**NIFI-9327**

Added timewindow query to QueryNiFiReportingTask and MetricsEventReportingTask.

**NIFI-9316**

Registry Sort by label should be 'Last Updated (newest)' not 'Newest (update)'.

**NIFI-9293**

Ensure that we properly set the scheduled flag in the LifecycleState when stopping processors. Added system test to verify that @OnScheduled, onTrigger, @OnUnscheduled, @OnStopped are all called and in the expected order.

**NIFI-9286**

- JOLT Expression Language Fixes NIFI-6213 and adds in functionality to use expression language in class and module specification.
- Adding JOLT unit tests.
- Addressing PR feedback Fixes a problem with the scope of the EL for module directory.
- Alignment of JOLT processors.
- Fix checkstyle.

**NIFI-9281**

Enabled building on Java 17.

**NIFI-9233**

Improve reliability of system integration tests.

**NIFI-9227**

Run Once not working when scheduling strategy is CRON or Event driven.

**NIFI-9166**

Refactored nifi-standard-services to use JUnit 5.

**NIFI-9134**

nifi-metrics-reporting-bundle to use JUnit 5.

**NIFI-9133**

Refactored nifi-media-bundle to use JUnit 5.

**NIFI-9124**

Refactored nifi-jms-bundle to use JUnit 5.

**NIFI-9113**

Refactored nifi-grpc-bundle to use JUnit 5.

**NIFI-9103**

Refactored nifi-datadog-bundle to use JUnit 5.

**NIFI-9085**

Refactored the Elasticsearch bundle to use JUnit 5.

**NIFI-9072**

Improvements to ValidateXML.

**NIFI-9065**

Add support for OAuth2AccessTokenProvider in InvokeHTTP.

**NIFI-9064**

Support Oracle timestamp when 'Use Avro Logical Types' is true for ExecuteSQLRecord and QueryDatabaseTableRecord.

**NIFI-9058**

Corrected AttributesToJSON Core Attributes filtering.

**NIFI-8927**

Add option to start/stop all controllers.

**NIFI-8899**

Add NiFi Registry version information to the Registry under an "about" button.

**NIFI-8676**

Added 'Tracking Entities' listing strategy to 'ListS3' and 'ListGCSBucket'.

**NIFI-8549**

Upgraded MiNiFi sensitive properties algorithm.

**NIFI-8521**

Removed nifi-influxdb-nar package from nifi-assembly.

**NIFI-8492**

Addressed issue in DatabaseReader class of IPLookupService that was attempting to set values on the JSON returned by MaxMind. Instead of modifying the object directly, we should use an Injectable in the Reader so that the value read will have the appropriate values but we don't need to modify those objects returned by MaxMind. Similar solution of NIFI-5814.

**NIFI-8209**

- Added Neo4J 4.X and 3.X clients by splitting the current controller service along release lines. This was necessary because Neo4J broke compatibility in their client drivers for Java between 3.X and 4.X at the Java API level.
- Updated module name.
- Updated parent module.
- Renamed a few misnamed modules.
- Updated 1.15.0-SNAPSHOT references in cypher v3 package.
- Updated neo4j 4.x driver.

**NIFI-8040**

When changing version of a flow, stop processors that have a state of Starting in addition to those with a state of Running.

**NIFI-7865**

amqp$header is splitted in the wrong way for "," and "}".

**NIFI-7840**

Upgrade to Groovy 3.0.8 and Spock 2.1.

**NIFI-7835**

Added authenticated SOCKS proxy support for SFTP.

**NIFI-7333**

Added OIDC trust store strategy property.

**NIFI-7192**

Added systemd reload to nifi.sh install on systemd servers.

**NIFI-6871**

Added HikariCPConnectionPool controller service.

**NIFI-6740**

Add configuration options to specify NiFi/Bootstrap communication ports.

**NIFI-6699**

Corrected SFTP symbolic link handling (#5744).

**NIFI-6390 and NIFI-1825**

When we write to a FlowFile and that results in a 0-byte FlowFile, remove the content claim altogether. This is more efficient to process, but far more importantly it prevents a 0-byte FlowFile from holding content in the Content Repository. Also fixed issue in which a Provenance Event cannot be replayed if there is no ContentClaim.

**NIFI-6266**

Corrected proxy FTP connect handling.

**NIFI-6047**

- Cleaned up code to allow tests to run against 1.13.0-snapshot Removed DMC.
- Started integrating changes from NIFI-6014.
- Added DMC tests.
- Added cache identifier recordpath test.
- Added additional details.
- Removed old additional details.
- Made some changes requested in a follow up review.
- Finished updates First round of code review cleanup Latest Removed EL from the dynamic properties. Finished code review requested refactoring. Checkstyle fix. Removed a Java 11 API.
- Renamed processor to DeduplicateRecord.

## Fixed Issues in Streams Messaging

Review the list of Streams Messaging issues that are resolved in Cloudera DataFlow for Data Hub 7.2.15.

### Kafka

**CDPD-29058: Migrate to log4j2 due to log4j1 end of life**

Kafka is migrated and uses log4j2 as a logging library. Additionally, log4j1 dependencies are removed with the exception of the Log4jAppender. Although the appender remains available, Cloudera recommends that you use the log4j2 implementation of the appender that is available in the log4j2 project.

**CDPD-29307: Kafka keystore and truststore type is not configured for Cruise Control metrics reporter**

The keystore and truststore types are now correctly supported by the Cruise Control metrics reporter in the Kafka broker.

**OPSAPS-62548: TopicMetrics get deleted from Cloudera Manager during restart or Kafka partition reassignment**

KafkaTopicMetrics are no longer deleted from the ServiceMonitor's Time-series database during a Kafka restart or a partition leader change.

### Schema Registry

**CDPD-35983: Unique constraint violation on load balanced Schema Registry cluster startup**

A concurrency issue in a multi-node Schema Registry setup is fixed where more nodes tried to initialize database state at the same time causing some of them to fail.

**CDPD-35469: Schema Registry responds with Internal Server Error when adding more schemas than defined in offset range**

Schema Registry responds with HTTP 409 response instead of HTTP 500 response while trying to add more schemas than defined in offset range.

**CDPD-33908: Remove or Upgrade Spring framework to 5.3.14+/5.2.19 due to CVE-2021-22060**

Removed Spring dependencies from Schema Registry because they were not used at all.

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

Fixed v009__create_registry_audit.sql to have create index refer to the lower case "atlas_events" object (the table). Made the script re-runnable since the table was already created where the script had already run.

**CDPD-31881: Schema Registry L1 test fails with socket timeout**

When more than one instance of Schema Registry is running on the same DB, "concurrent update" exceptions might have appeared in the Schema Registry log regarding changes to be sent to Atlas.

## Streams Messaging Manager

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

Fixed SMM REST throwing an internal server error when custom timestamps are provided while calling the "/api/v2{or v1}/admin/replication-stats" endpoint, or when a custom time period is provided on the ProducerDetail page in SMM UI.

**CDPD-33011: Selecting a consumer with no producers should show 0 producers in the filter panel**

On the overview page in the filter panel, when a consumer is selected that has no producers associated, the number of producers will be shown to be 0 of T, where T is the total number of producers.

**CDPD-32936: Selecting a producer with no consumers should show 0 consumers in the filter panel**

On the overview page in the filter panel, when a producer is selected that has no consumers associated, the number of consumers will be shown to be 0 of T, where T is the total number of consumers.

**CDPD-29403: When editing the alert, the topic can be chosen for the replication status**

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

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

The Kafka Connect tab is now correctly displayed if Kafka Connect is provisioned on the cluster.

**OPSAPS-62548: TopicMetrics get deleted from CM during restart or kafka partition reassignment**

KafkaTopicMetrics accidentally gets deleted from ServiceMonitor's Timeseries database during a Kafka restart or partition leader change.

## Streams Replication Manager

**CDPD-31745: SRM Control fails to configure internal topic when target is earlier than Kafka 2.3**

SRM now creates all internal topics explicitly. SRM also verifies the essential configurations of internal topics at startup, and fails if the topics do not meet the required configurations.

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

SRM Service Basic Authentication would not work with the default, random generated password. SRM Service Basic Authentication default password is now identical on all SRM Service role instances.

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

SRM uses the correct ssl.keystore.key configuration when a Kafka External Account specifies the keystore.

### Cruise Control

**Support added for keystore and truststore types other than JKS**

You are able to configure the keystore and truststore in Kafka brokers for Cruise Control Metrics Reporter. Previously, only the JKS type was supported for the SSL keystore and truststore.

**Migrating Cruise Control to Log4j2**

You are able to configure the keystore and truststore in Kafka brokers for Cruise Control Metrics Reporter. Previously, only the JKS type was supported for the SSL keystore and truststore.

**Cruise Control fails to start after upgrade with Rack Aware Goal configured**

You are able to configure the keystore and truststore in Kafka brokers for Cruise Control Metrics Reporter. Previously, only the JKS type was supported for the SSL keystore and truststore.

### Technical Service Bulletins

**7.2.15.4**

**TSB 2022-614: Kafka policy "user auto-creation" does not work in Ranger in CDP Public Cloud 7.2.15**

For the latest update on this issue see the corresponding Knowledge article: TSB 2022-614: Kafka policy "user auto-creation" does not work in Ranger in CDP Public Cloud 7.2.15

## Fixed Issues in Streaming Analytics

Review the list of Streaming Analytics issues that are resolved in Cloudera DataFlow for Data Hub 7.2.15.

### 7.2.15.6

**CSA-4127: Upgrading Streaming Analytics clusters**

The issue regarding the upgrade of Streaming Analytics clusters is fixed.

### 7.2.15.3

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

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

### 7.2.15.2

**CSA-2729: DLQ topic is filled with unexpected results**

The issue regarding not expected results written to the Dead Letter Queue (DLQ) topic is fixed.

**CSA-2797: Materialized View breaks when reloading stopped SQL job**

The issue about Materialized View failing when reloading a stopped SQL job has been fixed.

**CSA-3308: Changing Primary Key for Materialized View without recreating table causes job failure**

The issue regarding job failure when changing the primary key for Materialized View has been fixed.

**CSA-3464: Overwriting new changes for SQL Job with edit**

The issue about overwriting the changes of a SQL job when editing it on SQL Jobs tab is fixed.

**CSA-3537: Catalogs are deleted after Streaming SQL Engine restart**

The issue regarding catalogs being deleted after restarting the Streaming SQL Engine is fixed.

### 7.2.15

**CSA-2547: Vulnerability issue for user impersonation**

The vulnerability issue of using the doAs=other_user parameters is fixed. Users cannot be impersonated when using SPNEGO authentication.

**CSA-2529: SQL query fails when consumer group is set for Kafka**

The issue regarding failing SQL queries when a consumer group is set for Kafka has been fixed.

**CSA-2559: Materialized View settings can be overwritten while running job**

The issue of Materialized View settings are overwritten when submitting a new job with the same name has been fixed.

# Fixed CVEs in Cloudera DataFlow for Data Hub 7.2.15

Review the list of CVEs that are resolved in Cloudera DataFlow for Data Hub 7.2.15.

## 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 7.2.15.

**CVE-2020-36518**

The vulnerable jackson-databind dependency allowed a Java stack overflow exception and denial of service through a large depth of nested objects.

**CVE-2021-42392**

Apache NiFi uses H2 database for storing various NiFi runtime details. H2 database had a critical vulnerability similar to Log4Shell that potentially allows JNDI remote codebase loading. In NiFi,

by default, console access to the database is restricted to local machine access only and remote access is disabled, which limits the severity of this vulnerability. More detailed information on the H2 vulnerability can be found in this blog post. Note that the fix for this CVE impacts the list of external databases Cloudera supports for the NiFi Registry instance. See the Support Matrix for more information.

**CVE-2022-26850**

When creating or updating credentials for single-user access, NiFi wrote a copy of the Login Identity Providers configuration to the operating system temporary directory. The Login Identity Providers configuration file contains the username and a bcrypt hash of the configured password. On most platforms, the operating system temporary directory has global read permissions. NiFi immediately moved the temporary file to the final configuration directory, which significantly limited the window of opportunity for access.

**CVE-2022-29265**

Multiple components in Apache NiFi versions 0.0.1 to 1.16.0 do not restrict XML External Entity references in the default configuration. The Standard Content Viewer service attempts to resolve XML External Entity references when viewing formatted XML files. The following Processors attempt to resolve XML External Entity references when configured with default property values:

- EvaluateXPath
- EvaluateXQuery
- ValidateXml

Apache NiFi flow configurations that include these processors are vulnerable to malicious XML documents that contain Document Type Declarations with XML External Entity references.

# Behavioral Changes in Cloudera DataFlow for Data Hub 7.2.15

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 7.2.15.

## Behavioral Changes in Streaming Analytics

Review the list of Streaming Analytics behavioral changes in Cloudera DataFlow for Data Hub 7.2.15.

### 7.2.15.2

**Summary:**

Configurable checkpointing and failure restart strategy for SSB jobs.

Previous behavior:

Checkpointing and failure restart strategy followed the default Flink setting. This meant that both features were disabled by default and could not be configured using dedicated settings on Streaming SQL Console, only with SET statements for a specific job.

New behavior:

Both checkpointing and failure restart strategy features are enabled by default, and can be configured through Streaming SQL Console under the Settings panel of a SQL job.

**Summary:**

Materialized View exception handling

Previous behavior:

When an error occurred writing to the PostgreSQL database behind Materialized Views, the exception was caught and ignored, and did not affect the SQL job.

New behavior:

When an error occurs writing to the PostgreSQL database behind Materialized Views, the exception is logged, there is a retry attempt to allow jobs to recover before the job fails. When the job fails, the restart strategy and checkpointing settings are used.