Cloudera DataFlow for Data Hub 7.2.17

Cloudera DataFlow for Data Hub Release Notes

Date published: 2019-12-16
Date modified: 2023-06-27

https://docs.cloudera.com/

Legal Notice

Contents

What's new in Cloudera DataFlow for Data Hub 7.2.17

Cloudera DataFlow for Data Hub 7.2.17 includes components for Flow Management, Edge 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 7.2.17.

Flow Management in CDP Public Cloud 7.2.17 is based on Apache NiFi 1.21.0 and includes significant improvements and fixes. Here are the most important new features and improvements:

Schema Registry

For new clusters, a Schema Registry instance is provisioned on the management node next to Cloudera Manager and NiFi Registry. This allows you to leverage the Schema Registry capabilities without the need for a Streams Messaging Data Hub cluster.

Cloudera Schema Registry Controller Service

As the Hortonworks Schema Registry Controller Service will be removed in NiFi 2.0, a new Cloudera Schema Registry Controller Service is now available and you are invited to switch to the new controller service to prepare for the NiFi 2.0 release. For new clusters, Cloudera provisions default instances of the controller service that are pre-configured with the Schema Registry instance deployed in the cluster.

Java version

Starting with CDP 7.2.17, Java 8 and 11 are installed on the nodes by default. While Java 8 remains the default, it is possible to switch to another Java version for existing clusters, or after an upgrade. For more information, see Changing the Java version of Flow Management Data Hub clusters.

When you create a new cluster through the UI, Java 8 is used as the default JDK. The ability to deploy a new Flow Management cluster with Java 11 is available only using the CLI. For more information, see Creating your cluster.

New components:

- ConvertProtobuf to convert Protobuf data into JSON
- Decrypt and Encode Content capabilities
- Modify Compression processor
- GCP Vision components for image processing
- AWS Polly, AWS Textract, AWS Transcribe, AWS Translate
- Asana source processor
- Jira source processor
- JSLT processor for easier transformations of JSON data
- gRPC components with custom schema handling
- Box, Dropbox, Google Drive sink processors
- Support for Apache IoTDB
- Salesforce sink processor
- Redis sink processor (Technical Preview)
- DeltaLake sink processor (Technical Preview)
- Trigger Hive Metastore Event processor to update the HMS after pushing data into the object store
- Validate JSON
- Verify Content MAC
- Generate Record processor – to better simulate data feeds

- PostgreSQL controller service – to connect to PostgreSQL databases without the need to deal with the JDBC driver

What's new in Edge Management [Technical Preview]

Learn about the Technical Preview for Light Duty Edge Flow Management cluster definitions introduced in Cloudera DataFlow (CDF) for Data Hub 7.2.17 in CDP Public Cloud.

7.2.17.2

Edge Flow Management cluster definitions have been updated with CEM 1.6.0 functionalities, providing an enhanced user experience. Edge Flow Manager now supports the deployment and automatic configuration of MiNiFi agents, including security settings.

For more information about how to support agent provisioning using the Remote Agent Deployer, see Enabling remote agent deployment in Edge Flow Manager.

7.2.17.0

The new templates include Edge Flow Manager (EFM), and they work out of the box in a secure environment on CDP Public Cloud. You can create light duty clusters that provide management, control, and monitoring capabilities for existing agents to collect data from edge devices and push information back to the edge.

For more information about edge management in CDP Public Cloud, see Cloudera Edge Management.

What's new in Streams Messaging

Learn about the new Streams Messaging features in Cloudera DataFlow for Data Hub 7.2.17.

Kafka

Rebase on Kafka 3.4.0

Kafka shipped with this version of Cloudera Runtime is based on Apache Kafka 3.4.0. For more information, see the following upstream resources:

Apache Kafka Notable Changes:

- 3.2.0
- 3.3.0 and 3.3.1
- 3.4.0

Apache Kafka Release Notes:

- 3.2.0
- 3.3.0
- 3.3.1
- 3.4.0

Kafka KRaft [TECHNICAL PREVIEW]

Apache Kafka Raft (KRaft) is a consensus protocol used for metadata management that was developed as a replacement for Apache ZooKeeper. Using KRaft for managing Kafka metadata instead of ZooKeeper offers various benefits including a simplified architecture and a reduced operational footprint.

Kafka KRaft in this release of Cloudera Runtime is in technical preview and does not support the following:

- Deployments with multiple log directories. This includes deployments that use JBOD for storage.

- Delegation token based authentication.
- Migrating an already running Kafka service from ZooKeeper to KRaft.
- Atlas Integration.

For a conceptual overview on KRaft, see Kafka KRaft. For more information on how to deploy a Streams Messaging Data Hub cluster that is running KRaft mode, see Setting up your Streams Messaging cluster.

SMT plugins for binary conversion

Two Cloudera developed Single Message Transforms (SMT) plugins are added. These are the ConvertToBytes and ConvertFromBytes plugins, which you can use to convert binary data to or from the Kafka Connect internal data format.

For more information, see the following resources:

- Single Message Transforms
- ConvertFromBytes
- ConvertToBytes

EOS for source connectors

Exactly-once semantics (EOS) support is added for Kafka Connect source connectors. For more information, see Configuring EOS for source connectors.

Rolling restart checks provide a high cluster health guarantees by default

The default value of the Cluster Health Guarantee During Rolling Restart property is changed from none to healthy partitions stay healthy. This property defines what type of checks are performed during a Rolling Restart on the restarted broker. Each setting guarantees a different level of cluster health during Rolling Restarts. With the none setting, no checks are performed. This means that in previous versions no guarantees were provided on cluster health by default.

The new default, healthy partitions stay healthy, ensures a high level of guarantees on cluster health. This setting ensures that no partitions go into an under-min-isr state when a broker is stopped. This is achieved by waiting before each broker is stopped so that all other brokers can catch up with all replicas that are in an at-min-isr state. Additionally, the setting ensures that the restarted broker is accepting requests on its service port before restarting the next broker. This setting ignores partitions which are already in an under-min-isr state. For more information, see Configuring EOS for source connectors.

LDAPS SSL configurations are inherited from the Kafka broker

The SSL configurations of LDAP over SSL (LDAPS) are inherited from the Kafka broker. Previously, the JDK default was used. If the JDK default certificate store contains certificates which were used to setup SSL connection to LDAP, it should be imported to the broker stores.

Aliases for Kafka CLI tools

Aliases are added for the kafka-storage.sh, kafka-cluster.sh, and kafka-features.sh command line tools. These tools can now be called globally with kafka-storage, kafka-cluster, and kafka-features.

Important: Not all tools are fully supported and their use is limited. For more information, see Unsupported command line tools .

Schema Registry

KafkaAvroSerializer and KafkaAvroDeserializer improvements

KafkaAvroSerializer and KafkaAvroDeserializer can now handle null values without Avro

The KafkaAvroSerializer and KafkaAvroDeserializer now support a configuration property called null.passthrough.enabled, which is false by default. If enabled, null data is handled as null, and no schema is sent to Schema Registry. This behavior enables client applications to write tombstone messages into compact topics. The KafkaAvroDeserializer also handles null values by returning null without any regards to the schema.

Support deserialization when the topic and schema names don't match

From now on, the KafkaAvroDeserializer uses the schema version's ID in the Avro byte stream to access the actual schema and disregards schema names.

Logical types conversion for the KafkaAvroSerliazier and KafkaAvroDesrializer

The KafkaAvroSerializer and KafkaAvroDeserializer can now properly handle and convert Avro logical types at a record level. This means that if you have a record that has a field with a built-in Avro logical type (for example a BigDecimal field with BYTES type and decimal logical type), you can now properly serialize the records. After deserialization, a GenericRecord is returned, including the typed BigDecimal field, instead of a ByteBuffer. Logical type conversion can be enabled using the logical.type.conversion.enabled property. This property is set to false by default for backward compatibility.

For more information, see the following resources:

- KafkaAvroDeserializer properties reference
- KafkaAvroSerializer properties reference

Principal mapping rules can be defined without quotes

The SSL Client Authentication Mapping Rules (schema.registry.ssl.principal.mapping.rules) property now accepts rules that are defined without quotes. As a result, when adding multiple rules, you no longer need to enclose each rule in quotes.

Remove modules section from registry.yaml

In previous versions, the registry.yaml configuration file contained a modules section. This section was used to list pluggable modules that extended Schema Registry's functionality. However, modules were never fully supported and have been removed in a previous release. The modules section in registry.yaml was kept for backwards compatibility. Starting with this version, the modules section is removed by default from registry.yaml.

Streams Messaging Manager

UI updates

The style of SMM UI is updated. This update includes various changes to the colors, fonts, and overall style of the UI. Additionally, the following functional changes and improvements are made:
Data Explorer

- The modal window that you use to view messages now includes a copy to clipboard button if the message you are viewing is long.
- A ↺ (Refresh) option is added next to the FROM OFFSET field. This option refreshes the partition offset range and fetches the latest messages.

Connector Configuration and Connector Settings pages

- A new option, Add, is added to the **Import a Connector config…** modal. This option enables you to import connector configuration properties without overriding existing properties.
- Property keys can now be filtered based on their group and importance.
- A Reset Filters option is added, this option resets all search filters.
- Three new actions are added that modify the configuration as a whole. The options are Remove all, Reset, and Export. These actions are available in a new Actions drop-down.
- The Import Connector Configuration… option is moved to the Actions drop-down and is renamed to Import.
- The **Deployment Status** modal now correctly displays the status of the deployment process.
- An error message is added that notifies you if validation errors are found for properties that are currently filtered.
- If available, the display names of configuration property keys are displayed above the property key.

Highly available Kafka Connect integration

SMM uses the Kafka Connect service role's REST URL to establish a connection with Connect and serve Connect metrics. Previously, even if your Connect deployment was highly available and had multiple service roles deployed, SMM could only be configured with a single connection URL. From now on, multiple URLs can be configured. If the Connect service role that SMM is connected to fails, SMM automatically connects to a different instance that is available.

As a result of this change, the Kafka Connect Host and Kafka Connect Port properties are replaced by the Kafka Connect Rest HostPorts property. If Kafka Connect Rest HostPorts is left empty (default), SMM is automatically configured with the host, port, and protocol of the Connect service role instances belonging to the Kafka service selected with the Kafka Service SMM property.

If you previously configured Kafka Connect Host and Kafka Connect Port, the values set in the properties are automatically migrated to Kafka Connect Rest HostPorts when you upgrade.

The SMM API now hides email notifier SMTP passwords in its response

Previously the /notifiers endpoint returned the full configuration of the notifier. In the case of email notifiers, the configuration included the password of the SMTP server. API responses from now on do not include the password. As result of this change, the PASSWORD field of existing email notifiers is left blank when you edit them. If you decide to edit the notifier you must reenter the password.

Streams Replication Manager

Improved SRM logging

SRM's logging capabilities are improved. From now on:

- Kafka clients created by SRM's internal connectors reference the replication flow they are a part of (KAFKA-14838 backport).
- SRM now includes references to the replication flow in the log context of its internal connectors.

These changes enable differentiation between the logs associated with each replication flow.

Cruise Control

Rebasing Cruise Control to 2.5.116

Cruise Control in Cloudera Runtime is rebased to the 2.5.116 version. For more information about the fixes and features in Cruise Control 2.5.116, see the Cruise Control Rebase Summary.

New endpoint for Cruise Control

The GET/kafkacruisecontrol/permissions endpoint is added to Cruise Control that lists the level permissions of the current user. In case authentication is not configured for a user, the GET call returns Unable to retrieve privilege information for an unsecure connection message.

What's new in Streaming Analytics

Learn about the new Streaming Analytics features in Cloudera DataFlow for Data Hub 7.2.17.

The following new features are introduced in Streaming Analytics CDF for Data Hub 7.2.17:

Rebase to Apache Flink 1.16

Apache Flink 1.16 is supported in the Streaming Analytics 7.2.17 cluster definition.

For more information on what is included in the Apache Flink 1.16 version, see the Apache Flink 1.16 Release Post and the Apache Flink 1.16 Release Notes

Iceberg V2 Support

Iceberg V2 is supported for Flink and SQL Stream Builder. The V2 support allows row-level updates and deletes in Iceberg tables.

For more information, see the Iceberg with Flink and Creating Iceberg tables in SSB documentation.

Built-in Data Visualization tool

The built-in data visualization tool in SQL Stream Builder (SSB) enables you to present the sampling data and the results of the Materialized View query using widgets on Streaming SQL Console. As the widgets are integrated into SSB, the visualization tool works out of the box without any dependencies, which offers easy access to the underlying, running jobs as data sources.

For more information, see the Using widgets for data visualization documentation.

Schema Handling

Cloudera Schema Registry catalog can now handle key schemas as well based on the provided key (and value) schema name suffix for custom naming convention.

For more information, see the Adding Catalogs documentation.

Component support in Cloudera DataFlow for Data Hub 7.2.17

Cloudera DataFlow for Data Hub 7.2.17 includes the following components.

Flow Management clusters

- Apache NiFi 1.21.0.2.2.7.0
- Apache NiFi Registry 1.21.0.2.2.7.0

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

Edge Management clusters

- Edge Flow Manager 1.5.1.0

Streams Messaging clusters

- Apache Kafka 3.4.0
- Schema Registry 0.10.0
- Streams Messaging Manager 2.3.0
- Streams Replication Manager 1.1.0
- Cruise Control 2.5.116

Streaming Analytics clusters

- Apache Flink 1.16

Supported NiFi extensions

Apache NiFi 1.21.0 ships with a set of processors, controller services, and reporting tasks, most of which are supported by Cloudera. Review the supported extensions and avoid using any unsupported extensions in your production environments.

Supported NiFi processors

This release ships with Apache NiFi 1.21.0 and includes a set of processors, most of which are supported by Cloudera. 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	GetGcpVisionAnnotateFilesOperationStatus	PutDynamoDBRecord
AttributesToJSON	GetGcpVisionAnnotateImagesOperationStatus	PutElasticsearchHttp1
Base64EncodeContent	GetHBase	PutElasticsearchHttpRecord
CalculateRecordStats	GetHDFS	PutElasticsearchJson
CaptureChangeMySQL	GetHDFSFileInfo	PutElasticsearchRecord
CompressContent12	GetHDFSSequenceFile	PutEmail1
ConnectWebSocket	GetHTMLElement	PutFile
ConsumeAMQP	GetHTTP	PutFTP
ConsumeAzureEventHub	GetHubSpot	PutGCSObject
ConsumeEWS	GetIgniteCache	PutGoogleDrive
ConsumeGCPubSub	GetJiraIssue	PutGridFS
ConsumeGCPubSubLite	GetJMSQueue	PutHBaseCell1
ConsumeJMS	GetJMSTopic	PutHBaseJSON
ConsumeKafka_1_0	GetMongoRecord	PutHBaseRecord
ConsumeKafka_2_0	GetSFTP	PutHDFS
ConsumeKafka_2_6	GetShopify	PutHive3QL
ConsumeKafka2CDP	GetSNMP	PutHive3Streaming
ConsumeKafka2RecordCDP	GetSnowflakeIngestStatus	PutHiveQL
ConsumeKafkaRecord_1_0	GetSolr	PutHiveStreaming
ConsumeKafkaRecord_2_0	GetSplunk	PutHTMLElement
ConsumeKafkaRecord_2_6	GetSQS	PutIceberg
ConsumeKinesisStream	GetTCP	PutInfluxDB
ConsumeMQTT1	GetTwitter	PutJMS
ConsumeTwitter	GetWorkdayReport	PutKinesisFirehose
ConsumeWindowsEventLog	GetZendesk	PutKinesisStream
ControlRate	HandleHttpRequest	PutKudu
ConvertAvroSchema	HandleHttpResponse	PutLambda
ConvertAvroToJSON	HashAttribute	PutMongoRecord
ConvertAvroToORC	HashContent	PutORC
ConvertAvroToParquet	IdentifyMimeType	PutParquet
ConvertCharacterSet	InvokeAWSGatewayApi	PutRecord
ConvertCSVToAvro	InvokeGRPC	PutRedisHashRecord (Technical Preview)
ConvertJSONToAvro	InvokeHTTP	PutRiemann
ConvertJSONToSQL	InvokeScriptedProcessor	PutS3Object
ConvertProtobuf	JoinEnrichment	PutSalesforceObject
ConvertRecord	JoltTransformJSON	PutSFTP
CreateHadoopSequenceFile	JoltTransformRecord	PutSmbFile

CryptographicHashAttribute	JSLTTransformJSON	PutSnowflakeInternalStage
CryptographicHashContent	JsonQueryElasticsearch	PutSNS
DecryptContent	ListAzureBlobStorage	PutSolrContentStream
DecryptContentCompatibility	ListAzureBlobStorage_v12	PutSolrRecord
DecryptContentPGP	ListAzureDataLakeStorage	PutSplunk
DeduplicateRecord	ListBoxFile	PutSplunkHTTP1
DeleteAzureBlobStorage	ListCDPObjectStore	PutSQL
DeleteAzureBlobStorage_v12	ListDatabaseTables	PutSQS
DeleteAzureDataLakeStorage	ListDropbox	PutSyslog
DeleteByQueryElasticsearch	ListenBeats	PutTCP
DeleteCDPObjectStore	ListenFTP	PutUDP
DeleteDynamoDB	ListenGRPC	PutWebSocket1
DeleteGCSObject	ListenGRPC	QueryAirtableTable
DeleteGridFS	ListenHTTP	QueryCassandra
DeleteHBaseCells	ListenRELP	QueryDatabaseTable
DeleteHBaseRow	ListenSyslog	QueryDatabaseTableRecord
DeleteHDFS	ListenTCP	QueryElasticsearchHttp
DeleteS3Object	ListenTCPRecord	QueryRecord
DeleteSQS	ListenTrapSNMP	QuerySalesforceObject
DetectDuplicate	ListenUDP	QuerySolr
DistributeLoad	ListenUDPRecord	QuerySplunkIndexingStatus
DuplicateFlowFile	ListenWebSocket	QueryWhois
EncodeContent	ListFile	ReplaceText1
EncryptContent2	ListFTP	ReplaceTextWithMapping
EncryptContentPGP	ListGCSBucket	ResizeImage
EnforceOrder	ListGoogleDrive	RetryFlowFile
EvaluateJsonPath	ListHDFS	RouteHL7
EvaluateXPath	ListS3	RouteOnAttribute
EvaluateXQuery	ListSFTP	RouteOnContent
ExecuteGroovyScript	ListSmb	RouteText
ExecuteInfluxDBQuery	LogAttribute	SampleRecord1
ExecuteProcess	LogMessage	ScanAccumulo
ExecuteScript	LookupAttribute	ScanAttribute
ExecuteSQL	LookupRecord	ScanContent
ExecuteSQLRecord	MergeContent1	ScanHBase
ExecuteStateless	MergeRecord	ScriptedFilterRecord
ExecuteStreamCommand	ModifyCompression	ScriptedPartitionRecord
ExtractAvroMetadata	ModifyHTMLElement	ScriptedTransformRecord
ExtractGrok	MonitorActivity	ScriptedValidateRecord
ExtractHL7Attributes	MoveAzureDataLakeStorage	ScrollElasticsearchHttp

ExtractImageMetadata	MoveHDFS	SearchElasticsearch
ExtractText	Notify	SegmentContent
FetchAzureBlobStorage	PaginatedJsonQueryElasticsearch	SelectHive3QL
FetchAzureBlobStorage_v12	ParseCEF	SelectHiveQL
FetchAzureDataLakeStorage	ParseEvtx	SendTrapSNMP
FetchBoxFile	ParseSyslog	SetSNMP
FetchCDPObjectStore	PartitionRecord	SignContentPGP
FetchDistributedMapCache	PostHTTP	SplitAvro1
FetchDropbox	PublishAMQP1	SplitContent1
FetchElasticsearchHttp	PublishGCPubSub1	SplitJson1
FetchFile	PublishGCPubSubLite	SplitRecord
FetchFTP	PublishJMS1	SplitText1
FetchGCSObject	PublishKafka_1_0	SplitXml1
FetchGoogleDrive	PublishKafka_2_0	StartAwsPollyJob
FetchGridFS	PublishKafka_2_6	StartAwsTextractJob
FetchHBaseRow	PublishKafka2CDP	StartAwsTranscribeJob
FetchHDFS	PublishKafka2RecordCDP	StartAwsTranslateJob
FetchParquet	PublishKafkaRecord_1_0	StartGcpVisionAnnotateFilesOperation
FetchS3Object	PublishKafkaRecord_2_0	StartGcpVisionAnnotateImagesOperation
FetchSFTP	PublishKafkaRecord_2_6	StartSnowflakeIngest
FetchSmb	PublishMQTT1	TagS3Object
FlattenJson	PutAccumuloRecord	TailFile
ForkEnrichment	PutAzureBlobStorage	TransformXml
ForkRecord	PutAzureBlobStorage_v12	TriggerHiveMetaStoreEvent
GenerateFlowFile	PutAzureCosmosDBRecord1	UnpackContent
GenerateRecord	PutAzureDataLakeStorage	UpdateAttribute
GenerateTableFetch	PutAzureEventHub1	UpdateByQueryElasticsearch
GeoEnrichIP	PutAzureQueueStorage	UpdateCounter
GeoEnrichIPRecord	PutBigQuery	UpdateDatabaseTable
GeohashRecord	PutBigQueryBatch	UpdateDeltaLakeTable (Technical Preview)
GetAsanaObject	PutBigQueryStreaming1	UpdateHive3Table
GetAwsPollyJobStatus	PutBoxFile	UpdateHiveTable
GetAwsTextractJobStatus	PutCassandraQL1	UpdateRecord
GetAwsTranscribeJobStatus	PutCassandraRecord	ValidateCsv
GetAwsTranslateJobStatus	PutCDPObjectStore	ValidateJson
GetAzureEventHub	PutCloudWatchMetric	ValidateRecord
GetAzureQueueStorage	PutCouchbaseKey1	ValidateXml
GetCouchbaseKey1	PutDatabaseRecord	VerifyContentMAC
GetElasticsearch	PutDistributedMapCache	VerifyContentPGP
GetFile	PutDropbox	Wait

GetFTP	PutDynamoDB1	YandexTranslate

Footnotes

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

Supported NiFi controller services

This release ships with Apache NiFi 1.21.0 and includes a set of Controller Services, most of which are supported by Cloudera Support. You should be familiar with the available supported Controller Services, and avoid using any unsupported Controller Services in production environments.

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

AccumuloService	IPFIXReader
ActionHandlerLookup (Technical Preview)	IPLookupService
ADLSCredentialsControllerService	JASN1Reader
ADLSIDBrokerCloudCredentialsProviderControllerService	JMSConnectionFactoryProvider
AlertHandler (Technical Preview)	JndiJmsConnectionFactoryProvider
AvroReader	JsonConfigBasedBoxClientService
AvroRecordSetWriter	JsonPathReader
AvroSchemaRegistry	JsonRecordSetWriter
AWSCredentialsProviderControllerService	JsonTreeReader
AWSIDBrokerCloudCredentialsProviderControllerService	KafkaRecordSink_1_0
AzureBlobIDBrokerCloudCredentialsProviderControllerService	KafkaRecordSink_2_0
AzureCosmosDBClientService	KafkaRecordSink_2_6
AzureEventHubRecordSink	KerberosKeytabUserService
AzureStorageCredentialsControllerService	KerberosPasswordUserService
AzureStorageCredentialsControllerService_v12	KerberosTicketCacheUserService
AzureStorageCredentialsControllerServiceLookup	KeytabCredentialsService
CassandraDistributedMapCache	KuduLookupService
CassandraSessionProvider	LoggingRecordSink
CdpCredentialsProviderControllerService	LogHandler (Technical Preview)
CdpOauth2AccessTokenProviderControllerService	MongoDBControllerService
CEFReader	MongoDBLookupService
CiscoEmblemSyslogMessageReader	ParquetReader
ClouderaSchemaRegistry	ParquetRecordSetWriter
ConfluentSchemaRegistry	PostgreSQLConnectionPool
CouchbaseClusterService	PrometheusRecordSink
CouchbaseKeyValueLookupService	ReaderLookup
CouchbaseMapCacheClient	RecordSetWriterLookup
CouchbaseRecordLookupService	RecordSinkHandler (Technical Preview)

CSVReader	RecordSinkServiceLookup
CSVRecordLookupService	RedisConnectionPoolService
CSVRecordSetWriter	RedisDistributedMapCacheClientService
DatabaseRecordLookupService	RestLookupService
DatabaseRecordSink	ScriptedActionHandler (Technical Preview)
DBCPConnectionPool	ScriptedLookupService
DBCPConnectionPoolLookup	ScriptedReader
DistributedMapCacheClientService	ScriptedRecordSetWriter
DistributedMapCacheLookupService	ScriptedRecordSink
DistributedMapCacheServer	ScriptedRulesEngine (Technical Preview)
DistributedSetCacheClientService	SimpleDatabaseLookupService
DistributedSetCacheServer	SimpleKeyValueLookupService
EasyRulesEngineProvider (Technical Preview)	SimpleScriptedLookupService
EasyRulesEngineService (Technical Preview)	SiteToSiteReportingRecordSink
ElasticSearchClientServiceImpl	SmbjClientProviderService
ElasticSearchLookupService	SnowflakeComputingConnectionPool
ElasticSearchStringLookupService	StandardAsanaClientProviderService
EmailRecordSink	StandardAzureCredentialsControllerService
EmbeddedHazelcastCacheManager	StandardDropboxCredentialService
ExpressionHandler (Technical Preview)	StandardHashiCorpVaultClientService
ExternalHazelcastCacheManager	StandardHttpContextMap
FreeFormTextRecordSetWriter	StandardOauth2AccessTokenProvider
GCPCredentialsControllerService	StandardPGPPrivateKeyService
GrokReader	StandardPGPPublicKeyService
HadoopCatalogService	StandardPrivateKeyService
HadoopDBCPConnectionPool	StandardProxyConfigurationService
HazelcastMapCacheClient	StandardRestrictedSSLContextService
HBase_1_1_2_ClientMapCacheService	StandardS3EncryptionService
HBase_1_1_2_ClientService	StandardSnowflakeIngestManagerProviderService
HBase_1_1_2_ListLookupService	StandardSSLContextService
HBase_1_1_2_RecordLookupService	StandardWebClientServiceProvider
HBase_2_ClientMapCacheService	Syslog5424Reader
HBase_2_ClientService	SyslogReader
HBase_2_RecordLookupService	UDPEventRecordSink
Hive3ConnectionPool	VolatileSchemaCache
HiveCatalogService	WindowsEventLogReader
HiveConnectionPool	XMLReader
HortonworksSchemaRegistry	XMLRecordSetWriter

Supported NiFi reporting tasks

This release ships with Apache NiFi 1.21.0 and includes a set of reporting tasks, most of which are supported by Cloudera. 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 1.21.0 and includes a set of components built, maintained and supported by Cloudera partners. You should reach out directly to these partners in case you need assistance.

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

Processors supported by partners

- ConsumePulsar (v1.18.0)
- ConsumePulsarRecord (v1.18.0)
- PublishPulsar (v1.18.0)
- PublishPulsarRecord (v1.18.0)

Controller Services supported by partners

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

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

Unsupported features in Cloudera DataFlow for Data Hub 7.2.17

Some features exist within Cloudera DataFlow for Data Hub 7.2.17 components, but are not supported by Cloudera.

Unsupported Flow Management features

Some Flow Management features exist in Cloudera DataFlow for Data Hub 7.2.17, but are not supported by Cloudera.

NiFi

Technical preview features

The following features are available in Flow Management in Cloudera DataFlow for Data Hub 7.2.17 but are not ready for production deployment. Cloudera encourages you to explore these technical preview features in non-production environments and provide feedback on your experiences through the Cloudera Community Forums.

- The following rules engine, handler, and catalog controller services:
  - ActionHandlerLookup
  - AlertHandler
  - EasyRulesEngineProvider
  - EasyRulesEngineService
  - ExpressionHandler
  - LogHandler
  - RecordSinkHandler
  - ScriptedActionHandler
  - ScriptedRulesEngine
  - ScriptedRulesEngine
  - Processors:
    - PutRedisHashRecord
    - UpdateDeltaLakeTable

NiFi Registry

There are no unsupported features in this release.

Unsupported Edge Management features [Technical Preview]

See the unsupported features listed in the Cloudera Edge Management documentation.

Unsupported Streams Messaging features

Some Streams Messaging features exist in Cloudera DataFlow for Data Hub 7.2.17, 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.
- Kafka KRaft in this release of Cloudera Runtime is in technical preview and does not support the following:
  - Deployments with multiple log directories. This includes deployments that use JBOD for storage.
  - Delegation token based authentication.
  - Migrating an already running Kafka service from ZooKeeper to KRaft.
  - Atlas Integration.

Schema Registry

There are no updates for this release.

Streams Messaging Manager

There are no updates for this release.

Streams Replication Manager

There are no updates for this release.

Cruise Control

There are no updates for this release.

Related Information

Cloudera Community Forum
Setting up your Streams Messaging cluster

Unsupported Streaming Analytics features

Some Streaming Analytic features exist in Cloudera DataFlow for Data Hub 7.2.17, 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
- RAZ-enabled GCP environment

- 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 7.2.17

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera DataFlow for Data Hub 7.2.17.

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:

NiFi UI becomes unavailable when a bulletin is generated at Flow Controller level

When a bulletin is generated at the Flow Controller level, such as for reporting tasks or a disconnected node, the NiFi UI becomes unavailable and gets stuck with a loading spinner. The UI remains inaccessible until the bulletin at the Flow Controller level is cleared (a bulletin persists for 5 minutes after being generated).

The issue is a result of NIFI-11433 and is fixed in NIFI-11560. The fix is not included in this release.

To address the issue until CDP 7.2.17.100 becomes available, upgrade your Data Hub cluster using the CDP CLI and utilize an image containing the fix. For instructions on how to use the `upgrade-cluster` CLI command, see https://cloudera.github.io/cdp-dev-docs/cli-docs/datahub/upgrade-cluster.html.

The image IDs are:

- AWS: 11be770b-27d2-4818-b0a5-6ac79e77e971
- Azure: eff93217-445b-48f6-8f36-01cdb3df01ce
- GCP: 56200b06-b5d6-4eee-bba7-17bd5d4b9121

Incomplete Ranger policy for NiFi metrics in Cloudera Manager

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

KafkaRecordSink puts multiple records in one message

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

For more information, see NIFI-8326.

There is no workaround for this issue.

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

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

Update the configuration of the ReportLineageToAtlas reporting task:

1. Open the Global menu on the NiFi UI.

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

InferAvroSchema may fail when inferring schema for JSON data

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

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

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

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

Known issues in Edge Management [Technical Preview]

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

CEM-2563: EFM behind proxy - support for generated download URLs

As a DataHub deployment EFM is behind proxy (Knox) and there are URL rewrite rules in place so user facing URLs that are generated by EFM are broken. Currently, this only affects the debug bundle download.

The generated download debug bundle URL is similar to this: https://log-coll-2-management0.cysec-en.a465-9q4k.cloudera.site/efm/api/transfer/3267dcbb-9341-4f30-bbb9-7b2a93de2725

It needs to be adjusted to be proxy aware by including the deployment name /cdp-proxy-api/efm-api mapping. The correct URL should be something like this: https://log-coll-2-management0.cysec-en.a465-9q4k.cloudera.site/log-coll-2/cdp-proxy-api/efm-api/transfer/3267dcbb-9341-4f30-bbb9-7b2a93de2725. Using the adjusted URL downloads the debug bundle.

For further known issues, see the Cloudera Edge Management documentation.

Known issues in Streams Messaging

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

Kafka

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

Known Issues

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

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

None

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

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

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

KAFKA-2561: Performance degradation when SSL Is enabled

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

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

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

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

None

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

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

None.

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

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

```
org.apache.kafka.common.KafkaException: Cannot execute transacti
onal method because we are in an error state
    at org.apache.kafka.clients.producer.internals.TransactionM
anager.maybeFailWithError(TransactionManager.java:1125)
    at org.apache.kafka.clients.producer.internals.TransactionMa
nager.maybeAddPartition(TransactionManager.java:442)
    at org.apache.kafka.clients.producer.KafkaProducer.doSend(K
afkaProducer.java:1000)
    at org.apache.kafka.clients.producer.KafkaProducer.send(Kafk
aProducer.java:914)
    at org.apache.kafka.clients.producer.KafkaProducer.send(Kafk
aProducer.java:800)
    .
    .
    .
Caused by: org.apache.kafka.common.errors.ClusterAuthorization
Exception: Cluster authorization failed.
```

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

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

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

CDPD-49304: AvroConverter does not support composite default values

AvroConverter cannot handle schemas containing a STRUCT type default value.

None.

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

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

None.

Limitations

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

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

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

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

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

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

      The Kafka service name is the value of $SERVICENAME.
2. Turn off the collection of partition level metrics:
   a. Go to 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.

Schema Registry

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

CDPD-49304: AvroConverter does not support composite default values

AvroConverter cannot handle schemas containing a STRUCT type default value.

None.

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-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-59015: Schema Registry does not create new versions of schemas even if the schema is changed

Schema Registry uses a schema fingerprinting mechanism to differentiate between schemas. However, fingerprinting does not take into consideration the schema attributes of the field type. As a result, if you have two schemas where the only difference is that one has type attributes defined and the other does not, they will be considered identical by Schema Registry. For example, the following schemas are considered identical:

```
#Schema V1
{"type":"record","name":"schema_name","namespace":"ns","fields":[
{"name":"local_timestamp_micros_long","type":"long"}]}

#Schema V2
{"type":"record","name":"schema_name","namespace":"ns","fields":[
{"name":"local_timestamp_micros_long","type":{"type":"long","log
icalType":"local-timestamp-micros"}}]}
```

Notice that the only difference is that in the second schema, the local_timestamp_micros_long field has a logical type specified. In cases like this, the new version of the schema is not created, the initial version is used. This is true even if the data that is being produced has a new schema version. The ID of the first schema version is used and is put in the serialized record. The new schema version is not created.

This issue is common when using change data capture (CDC) connectors like the Debezium connectors. This is because CDC connectors create schemas with the logical type decimal based on the column type in the database schema. For example:

```
{"type":"record","name":"schema_name","namespace":"ns","fields":
[{"name":"database_column","type":{"type":"bytes","logicalType":
"decimal","precision":64,"scale":0}}]}
```

If the database schema changes (for example, the column type), it is possible that only scale changes, which is a schema attribute.

```
{"type":"record","name":"schema_name","namespace":"ns","fields":
[{"name":"database_column","type":{"type":"bytes","logicalType":
"decimal","precision":64,"scale":1}}]}
```

In this case, even though scale changed to 1, the first version of the schema is used where scale is 0. As a result, the data is consumed with the wrong scale.

Avoid using logical types or other attributes. Alternatively, ensure that there are no changes in the logical types or other attributes between schema versions.

OPSAPS-68708: Schema Registry might fail to start if a load balancer address is specified in Ranger

Schema Registry does not start if the address specified in the Load Balancer Address Ranger property does not end with a trailing slash (/).

Set the value of the RANGER_REST_URL Schema Registry environment variable to an address that includes a trailing slash.

1. In Cloudera Manager, select the Schema Registry service.
2. Go to Configuration.
3. Find the Schema Registry Server Environment Advanced Configuration Snippet (Safety Valve) property and add the following:

   ```
   Key: RANGER_REST_URL
   Value: [***RANGER REST API URL***]
   ```

   Replace *[***RANGER REST API URL***]* with an address that can be used by Schema Registry to access Ranger. Ensure that the address ends with a trailing slash. For example: https://ranger-1.cloudera.com:6182/
4. Restart the Schema Registry service.

CDPD-58949: Schemas are de-duplicated on import

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

None.

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

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

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

None.

Streams Messaging Manager

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

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

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

None.

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

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

None.

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.

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

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

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

Limitations

SRM cannot replicate Ranger authorization policies to or from Kafka clusters

Due to a limitation in the Kafka-Ranger plugin, SRM cannot replicate Ranger policies to or from clusters that are configured to use Ranger for authorization. If you are using SRM to replicate data to or from a cluster that uses Ranger, disable authorization policy synchronization in SRM. This can be achieved by clearing the Sync Topic Acls Enabled (sync.topic.acls.enabled) checkbox.

Cruise Control

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

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

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

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

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

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

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

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

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

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

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

If a Cruise Control rebalance is initiated with the rebalance_disk parameter and Cruise Control is configured to fetch metrics from Cloudera Manager (Metric Reporter is set to CM metrics reporter),

Cruise Control stops collecting metrics from the partitions that are moved. This is because Cloudera Manager does not collect metrics from moved partitions due to an issue in Kafka (KAFKA-10320).

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

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

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

OPSAPS-68148: Cruise Control rack aware goal upgrade handler

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

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

Known issues in Streaming Analytics

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

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

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

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

```
...
<dependency>
 <groupId>org.apache.flink</groupId>
 <artifactId>flink-connector-kafka</artifactId>
 <version>${flink.version}</version>
 <exclusions>
  <exclusion>
   <groupId>org.apache.kafka</groupId>
   <artifactId>kafka-clients</artifactId>
  </exclusion>
 </exclusions>
</dependency>
<dependency>
 <groupId>org.apache.kafka</groupId>
 <artifactId>kafka-clients</artifactId>
 <version>3.1.2.7.2.17.0-334</version>
</dependency>
...
```

SQL Stream Builder

CSA-5138 - SQL job submissions with UDF JARs fail when checkpointing is enabled

Due to the handling of ClassLoaders for custom JARs, uploading any Java UDFs with checkpoints enabled will cause the SQL job to fail with the following error:

```
ERROR com.cloudera.ssb.sqlio.service.SqlExecutorService: Error w
hile submitting streaming job
org.apache.flink.util.FlinkRuntimeException: org.apache.flink.
api.common.InvalidProgramException: Table program cannot be comp
iled.
```

Once the SQL job fails, the session on Streaming SQL Console must be reset before resubmitting the job without checkpointing.

None

CSA-4858 - Kerberos encryption type detection does not always work correctly for SSB

SSB detects no supported encryption types even though there is a list of allowed encryption types in the krb5.conf file. This causes an error when generating keytabs from the principal and password pair.

1. Run ktutil on your cluster.
2. Change the configuration with the following commands:

```
addent -password -p <username> -k 1 -e aes256-cts
wkt /tmp/new_keytab.keytab
```

3. Upload the new keytab on Streaming SQL Console.

CSA-4800 - ToString of Job can cause stack overflow

The jobLogItems can cause stack overflow errors when toString is called.

None

CSA-4799 - Table Metadata is not saved when job is run via sql/execute

The SqlExecutorService.persistIfNeededAndExecute does not save the table metadata before executing the SQL job, therefore the data cleaner of the Materialized View Engine does not clean up the data based on the retention settings.

None

CSA-4699 - Keytab upload starts failing in SSB after some time, requiring a restart

The /tmp/ssb_keytab_work_dir is removed after a period of time and SSB can no longer create keytabs as the directory does not exist anymore.

Restart the SSB service to recreate the directory for the keytabs.

CSA-4650: Inconsistent sidebar collapse behavior

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

None

Limitations when configuring widgets

The following widget configuration optionns are not available for certain widgets on Streaming SQL Console:

- Gauge visualization type: Expand on hover, Unit
- Donut visualization type: Expand on hover, Title
- Pie visualization type: Expand on hover

None

Flink

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 7.2.17

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 7.2.17.

Fixed issues in Flow Management

Review the list of Flow Management issues that are resolved in Cloudera DataFlow for Data Hub 7.2.17.

CFM 2.2.7

CFM 2.2.7 is based on Apache NiFi 1.21.0. It includes all fixed issues of this Apache NiFi release, as well as the following additional fixes:

NIFI-11614

Improved Validation for JndiJmsConnectionFactoryProvider

NIFI-11493

Defaulted dynamically modified classpath fix

NIFI-11149

Added PutRedisHashRecord Processor

NIFI-11466

Added ModifyCompression Processor

NIFI-11034

Restore image viewer to nifi-assembly

NIFI-11502

Upgrade json-path to 2.8.0

NIFI-10865

Allow RecordPath's unescapeJson to convert de-serialised JSON Objects into Records

NIFI-11501

Upgraded Hive MQTT client from 1.3.0 to 1.3.1

NIFI-11500

Upgraded Jackson BOM from 2.14.2 to 2.15.0

NIFI-11498

Upgraded Netty from 4.1.91 to 4.1.92

NIFI-11497

Upgraded snowflake-ingest-sdk from 1.1.1 to 1.1.3

NIFI-11495

Upgraded Iceberg from 1.2.0 to 1.2.1

NIFI-11494

Upgraded JUnit from 5.9.2 to 5.9.3

NIFI-11473

Flow version change should not stop a component when only position is changed

NIFI-11483

Correctly use DescribedValue for JsonQueryElasticsearch

NIFI-11067

Delete Property History when changing Sensitive status

NIFI-11224

Refactor and FF attribute support in WHERE in QuerySalesforceObject.

NIFI-11433

Use .add() for set instead of .push()

NIFI-11111

Add option to output Elasticsearch error responses as FlowFile to PutElasticsearchJson and PutElasticsearchRecord

NIFI-11433

Update angular, d3, moment, slickgrid, and jquery dependencies

NIFI-11472

Check working directory after making directory in PutFTP

NIFI-11476

Populate resource consumitions

NIFI-11479

Upgraded JanusGraph from 0.5.3 to 0.6.3

NIFI-11478

Upgraded Spring Framework from 5.3.26 to 5.3.27

NIFI-11438

Set standard OpenID Connect Scopes

NIFI-11475

Fixed missing jackson-dataformat-xml in nifi-azure-services-api

NIFI-11402

PutBigQuery fix for case sensitivity and error handling

NIFI-11327

Add Export/Import All - NiFi CLI - NiFi Registry

NIFI-11461

Improve User and Group Tenants Search

NIFI-11387

Added Transport Type property to Azure Event Hub Components

NIFI-11255

Allowable value for 'Use s3.region Attribute'

NIFI-11435

NiFi CLI - add possibility to set 'Maximum Timer Driven Thread Count'

NIFI-11460

Switched to JLine 3.23.0 FileNameCompleter

NIFI-11386

Added Resource and Audience support to StandardOauth2AccessTokenProvider

NIFI-11458

Upgraded JNA from 5.12.1 to 5.13.0

NIFI-11456

Upgraded brotli4j from 1.8.0 to 1.11.0

NIFI-11457

Upgraded commons-configuration2 from 2.8.0 to 2.9.0

NIFI-11440

Speed up Iceberg Hive Metastore Tests

NIFI-11439

Corrected GCS Transit URI for custom Storage API URL

NIFI-11439

Added Storage API URL property to GCS Processors

NIFI-11437

Switched to StreamUtils.fillBuffer() for buffer, Improved EncryptContentPGP Content Type Detection

NIFI-11429

Upgrade Gremlin to 3.6.2

NIFI-11436

Fix NPE during updateFlow when called from a replace request for a PG that is not under version control

NIFI-10955

Added optional preprocessing to JASN1Reader

NIFI-11428

Upgraded Groovy to 3.0.17 and spock-core to 2.3

NIFI-11427

Upgraded Atlas from 2.2.0 to 2.3.0

NIFI-11426

Upgraded JLine from 3.22.0 to 3.23.0

NIFI-11366

Proxy aware C2 communication

NIFI-11407

Upgraded Azure BOM from 1.2.9 to 1.2.11

NIFI-11408

Enable user to disable gzip compression with PutGCSObject

NIFI-11417

Upgraded Caffeine from 2.8.1 to 2.9.3

NIFI-11421

Upgraded Parquet from 1.12.3 to 1.13.0

NIFI-11419

Upgraded MongoDB driver from 4.8.2 to 4.9.1

NIFI-11418

Upgraded bcrypt from 0.9.0 to 0.10.2

NIFI-11415

Upgraded Saxon-HE from 10.6 to 12.1

NIFI-11414

Upgraded snowflake-ingest-sdk from 1.0.3 to 1.1.1

NIFI-11413

Upgraded JUnit 5.9.1 to 5.9.2

NIFI-11412

Upgraded AspectJ Weaver to 1.9.6 for MiNiFi

NIFI-11410

Upgraded multiple integration dependencies

NIFI-11422

Fix stateless-assembly README typo

NIFI-11404

Upgraded AMQP Client from 5.16.0 to 5.17.0

NIFI-11403

Upgraded Calcite Core from 1.32.0 to 1.34.0

NIFI-11405

Upgraded multiple drivers and test dependencies

NIFI-11406

Upgraded Google GCP from 26.4.0 to 26.12.0

NIFI-11401

Upgraded snappy-java from 1.1.8.4 to 1.1.9.1

NIFI-11400

Upgraded Kotlin from 1.8.10 to 1.8.20

NIFI-11399

Upgraded Groovy from 3.0.14 to 3.0.17

NIFI-11398

Upgraded Apached HttpClient from 4.5.13 tp 4.5.14

NIFI-11397

Upgraded jsoup from 1.15.3 to 1.15.4

NIFI-11396

Upgraded AWS from 1.12.371 to 1.12.444

NIFI-11395

Upgraded Testcontainers from 1.17.6 to 1.18.0

NIFI-5642

QueryCassandra processor : output FlowFiles as soon fetch_size is reached

NIFI-11342

HDFS processors fail to get ClassloaderIsolationKey at startup

Fixed issues in Streams Messaging

Review the list of Streams Messaging issues that are resolved in Cloudera DataFlow for Data Hub 7.2.17.

Kafka

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

The Kafka-Atlas plugin now fully creates producer and consumer entities and does not generate incomplete ones.

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

The AvroConverter now propagates field default values to Connect schemas.

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

Kafka Connect does not fail on start retries when the Require Connectors To Override Kafka Client JAAS Configuration property is selected.

CDPD-53179: Amazon S3 Sink connector fails when buffer size is reached

The Amazon S3 Sink connector no longer fails when there is more than 5 MB (buffer size) of data available in a Kafka source topic and the connector receives more than 5 MB of data in a single poll.

Apache patch information

- KAFKA-14838: Add flow/connector/task/role information to MM2 Kafka client.id configs

Schema Registry

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

Schema Registry Amazon S3 JAR storage now functions correctly.

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

Schema Registry now evicts Kerberos user and group information from its cache with a configurable time.

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

The KafkaJsonSerializer and KafkaJsonDeserializer now properly translates null payloads as null.

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

The AvroConverter now propagates field default values to Connect schemas.

CDPD-48888: Schema Registry generates redundant schemas when byte[] with default field exists

Schema Registry's schema normalization and fingerprinting mechanism has been enhanced to properly handle default values for bytes data types.

CDPD-53380: Schema Registry Client should retry the request on Knox gateway errors

The Schema Registry Client will retry Knox gateway related failed requests as defined by the request retry configuration.

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 now visible in the Cloudera Schema Registry UI.

In addition, the `/api/v1/schemaregistry/search/schemas/aggregated` endpoint of the Cloudera Schema Registry API now correctly returns schemas created with the Confluent Schema Registry API.

Streams Messaging Manager

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

The **Consumer Group Profile** page now correctly shows the consumer instances on the right hand side. Previously the consumer groups were shown.

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

Using the search field on the Connect Cluster Profile tab no longer freezes the page.

CDPD-45406: The Connector Profile page of unassigned connectors is blank

The **Connector Profile** page of unassigned connectors are now correctly rendered and display that the connector is in an unassigned status.

CDPD-46073: Data Explorer loads indefinitely

The **Data Explorer** page no longer breaks if the partition parameter is manually removed from the URL.

CDPD-26633: The SMM API returns SMTP passwords of email notifiers in its response

The /notifiers endpoint of the SMM API no longer returns the SMTP password in its responses.

CDPD-49227: The Cluster Replications page crashes if the co-located cluster unknown to SRM

The **Cluster Replications** page is now correctly displayed even when the co-located Kafka cluster is unknown to SRM.

CDPD-56086: The Data Explorer modal displays the messages of the wrong topic

The **Data Explorer** modal that you open by clicking  on the **Topics** page now displays the messages of the selected topic.

CDPD-49696: Certain alerts may crash the Alerts page

Composite alerts with one of the conditions containing an assertion on cluster metrics no longer crashes the UI.

Streams Replication Manager

There are no fixed issues for Streams Replication Manager in Cloudera DataFlow for Data Hub 7.2.17.

Cruise Control

OPSAPS-66403: Cruise Control authenticated user are generated incorrectly

Cruise Control handles the different authentication levels for the users correctly. When users are added to higher authentication levels, the lower level permissions are also assigned. For example, an ADMIN level user automatically has USER and VIEWER permissions as well.

Fixed issues in Streaming Analytics

Review the list of Streaming Analytics issues that are resolved in Cloudera DataFlow for Data Hub 7.2.17.

7.2.17.4

CSA-5138 - SQL job submissions with UDF JARs fail when checkpointing is enabled

The issue regarding SQL job submission is fixed, uploading Java UDFs does not cause SQL job failure when checkpointing is enabled.

7.2.17.1

CSA-4779: Backporting Apache Flink 1.16.2 fixes

The fixes included in the Apache Flink 1.16.2 version is backported. For more information about the list of issues fixed in Apache Flink 1.16.2, see the Apache Flink 1.16 Release Announcement.

FLINK-30966: Flink SQL IF FUNCTION logic error

The fix regarding the logic error in FLINK-30966 is backported

7.2.17.0

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

The issue about creating a webhook table when the template is empty is fixed.

CSA-4333 - Use Kafka Timestamps switch reflects invalid value

The issue about incorrect status of the Kafka Timestamp configuration when viewing DDL is fixed.

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

The issue regarding the describe error for Virtual Tables using JSON and Avro schemas is fixed.

CSA-4400 - Cannot delete invalid catalog

The issue regarding invalid catalogs can be registered, but cannot be deleted is fixed.

CSA-4412 - Cannot delete mv endpoint when it contains dynamic parameters

The issue regarding Materialized View endpoint cannot be deleted when using dynamic parameters is fixed.

CSA-4425 - Password in Kafka Data Source can be revealed after save

The issue regarding viewing the password after saving the Kafka Data Source is fixed.

CSA-4426 - Kafka Data Source name accepts spaces

The issue about accepting Kafka Data Sources with spaces in their name is fixed.

CSA-4427 - State of Execute and Stop options in Job context menu do not correspond the Job state

The issue regarding the invalid property in the upsert Kafka table template is fixed.

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

The issue regarding the invalid property in the upsert Kafka table template is fixed.

CSA-4548 - Files cannot be uploaded through Swagger

The issue about the error when uploading files using REST API with Swagger is fixed.

CSA-4620 - Encryption for environment properties

The issue regarding encrypted environment properties is fixed.

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

The issue regarding the ignored parameters that added to the flink-yarn-session in command line is fixed.

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

The fix regarding the Calcite error in FLINK-18027 is backported.

Fixed CVEs in Cloudera DataFlow for Data Hub 7.2.17

Review the list of CVEs that are resolved in Cloudera DataFlow for Data Hub 7.2.17.

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) 2.2.7 in Data Hub in CDP Public Cloud 7.2.17.

CVE-2023-34212

Potential Deserialization of Untrusted Data with JNDI in JMS Components. The JndiJmsConnectionFactoryProvider controller service along with the ConsumeJMS and PublishJMS processors, in Apache NiFi from 1.8.0 through 1.21.0 allow an authenticated and authorized user to configure URL and library properties that enable deserialization of untrusted data from a remote location. The resolution validates the JNDI URL and restricts locations to a set of allowed schemes. For more information, see Behavioral changes.

Behavioral changes in Cloudera DataFlow for Data Hub 7.2.17

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 7.2.17.

Behavioral changes in Streams Messaging

Review the list of Streams Messaging behavioral changes in Cloudera DataFlow for Data Hub 7.2.17.

Kafka

Summary:

The Cluster Health Guarantee During Rolling Restart property is now set to healthy partitions s tay healthy. This change is done so that a higher level of cluster health guarantees are provided by default.

Previous behavior:

The default value of the Cluster Health Guarantee During Rolling Restart property was set to none.

New behavior:

The default value of the Cluster Health Guarantee During Rolling Restart property is set to healthy partitions stay healthy.

Schema Registry

There are no behavioral changes for Schema Registry in Cloudera DataFlow for Data Hub 7.2.17.

Streams Messaging Manager

There are no behavioral changes for Streams Messaging Manager in Cloudera DataFlow for Data Hub 7.2.17.

Streams Replication Manager

There are no behavioral changes for Streams Replication Manager in Cloudera DataFlow for Data Hub 7.2.17.

Cruise Control

There are no behavioral changes for Cruise Control in Cloudera DataFlow for Data Hub 7.2.17.

Behavioral changes in Streaming Analytics

Review the list of Streaming Analytics behavioral changes in Cloudera DataFlow for Data Hub 7.2.17.

SQL Stream Builder

Summary:

Dynamic MV parameters have different configuration steps

Previous behavior:

Dynamic parameters were specified after the Materialized View query is submitted on the created endpoint URL.

New behavior:

Dynamic and static parameters can be selected when configuring the Materialized View query.

Summary:

Stateless SSB API

Previous behavior:

Requests from the SSB API could only be used for the active project that was selected.

New behavior:

The new v2 SSB API is based on the project ID that is specified in the endpoint.

Behavioral changes in Flow Management

Review the list of Flow Management behavioral changes in Cloudera DataFlow for Data Hub 7.2.17.

Summary:

The Neo4JCypher3ClientService Controller Service has been completely removed in favor of the Neo4JCypherClientService controller service, which uses a more recent version of the underlying library.

Summary:

As part of NIFI-11614 and to ensure better security, some restrictions around the JndiJmsConnectionFactoryProvider controller service have been implemented.

New behavior:

The default validation for the JNDI Provider URL property only allows the following URL schemes:

- file
- jgroups
- ssl
- t3
- t3s
- tcp
- udp
- vm

If an additional URL scheme is required to interact with a specific JMS solution, a NiFi admin has to configure the following Java system property in the application bootstrap.conf file to override the default list: java.arg.jndiJmsUrlSchemesAllowed=-Dorg.apache.nifi.jms.cf.jndi.provider.url.schemes.allowed=ssl tcp

Note: The property must contain a space-separated list of URL schemes.