

Cloudera DataFlow for Data Hub 7.2.7

Cloudera DataFlow for Data Hub Release Notes

Date published: 2019-12-16

Date modified: 2021-02-09

CLUDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

What's New in Cloudera DataFlow for Data Hub 7.2.7.....	4
What's New in Flow Management.....	4
What's New in Streams Messaging.....	4
What's New in Streaming Analytics.....	5
Component Support in Cloudera DataFlow for Data Hub 7.2.7.....	5
Supported NiFi Extensions.....	5
Supported NiFi Processors.....	5
Supported NiFi Controller Services.....	7
Supported NiFi Reporting Tasks.....	8
Unsupported Features in Cloudera DataFlow for Data Hub 7.2.7.....	8
Unsupported Flow Management features.....	9
Unsupported Streams Messaging features.....	9
Unsupported Streaming Analytics features.....	9
Apache Patch Information in Cloudera DataFlow for Data Hub 7.2.7.....	10
NiFi patches.....	10
NiFi Registry patches.....	10
Known Issues In Cloudera DataFlow for Data Hub 7.2.7.....	10
Known Issues in Flow Management.....	10
Known Issues in Streams Messaging.....	11
Known Issues in Streaming Analytics.....	14
Fixed Issues in Cloudera DataFlow for Data Hub 7.2.7.....	14
Fixed Issues in Flow Management.....	14
Fixed Issues in Streams Messaging.....	15
Fixed Issues in Streaming Analytics.....	15

What's New in Cloudera DataFlow for Data Hub 7.2.7

Cloudera DataFlow for Data Hub 7.2.7 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 7.2.7.

The Flow Management clusters in CDP Public Cloud introduce the following new features:

- The `ConfluentSchemaRegistry` Controller Service now supports basic authentication.

[NIFI-6576](#)

- The `ReportLineageToAtlas` Reporting Task now supports `aws_s3_v2_directory`.

[NIFI-8131](#)

- External database support

Starting with CDP 7.2.7, the light and heavy duty templates instantiate an external database by default, (Amazon RDS, Azure SQL, Google Cloud SQL) to store data from Cloudera Manager, NiFi Registry, and Knox. The heavy duty template instantiates a database with high availability according to the options provided by the cloud provider.

This change is to provide better reliability of the Data Hub cluster in case of issue on the management node. This may cause a cost increase with your cloud provider. If you do not want this change, you can duplicate the template to create your own one and override the `externalDatabase` property with:

```
"externalDatabase" :  
{ "availabilityType": "NONE" }
```

Related Information

[Cloudera Support Portal](#)

What's New in Streams Messaging

Learn about the new Streams Messaging features in Cloudera DataFlow for Data Hub 7.2.7.

Kafka

There are no new features for Kafka in this release.

Schema Registry

The Schema Registry version is upgraded to 0.10.0

In the new version, the logic for High Availability (HA) was removed. Schema Registry can still work in HA/stateless mode behind a load balancer, but it cannot achieve HA on its own.

Streams Messaging Manager

There are no new features for Streams Messaging Manager in this release.

Streams Replication Manager

There are no new features for Streams Replication Manager in this release.

What's New in Streaming Analytics

Learn about the new Streaming Analytics features in Cloudera DataFlow for Data Hub 7.2.7.

There are no new features for Streaming Analytics in Cloudera DataFlow for Data Hub 7.2.7.

Component Support in Cloudera DataFlow for Data Hub 7.2.7

Cloudera DataFlow for Data Hub 7.2.7 includes the following components.

Flow Management clusters

- Apache NiFi 1.11.4
- Apache NiFi Registry 0.5.0

Streams Messaging clusters

- Apache Kafka 2.5.0
- Schema Registry 0.10.0
- Streams Messaging Manager 2.1.0
- Streams Replication Manager 1.0.0

Streaming Analytics clusters

- Apache Flink 1.10

Supported NiFi Extensions

Apache NiFi 1.11.4 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

Apache NiFi 1.11.4 ships with 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.

Supported processors

- AttributesToCSV
- AttributesToJSON
- Base64EncodeContent
- CalculateRecordStats
- CaptureChangeMySQL
- CompressContent
- ConnectWebSocket
- GeoEnrichIPRecord
- GetAzureEventHub
- GetAzureQueueStorage
- GetCouchbaseKey
- GetFile
- GetFTP
- GetHBase
- PutDistributedMapCache
- PutDruidRecord
- PutDynamoDB
- PutElasticsearch
- PutElasticsearch5
- PutElasticsearchHttp
- PutElasticsearchHttpRecord

- ConsumeAMQP
- ConsumeAzureEventHub
- ConsumeEWS
- ConsumeGCPubSub
- ConsumeJMS
- ConsumeKafka
- ConsumeKafka_0_10
- ConsumeKafka_1_0
- ConsumeKafka_2_0
- ConsumeKafkaRecord_0_10
- ConsumeKafkaRecord_1_0
- ConsumeKafkaRecord_2_0
- ConsumeMQTT
- ConsumeWindowsEventLog
- ControlRate
- ConvertAvroSchema
- ConvertAvroToJSON
- ConvertAvroToORC
- ConvertAvroToParquet
- ConvertCharacterSet
- ConvertCSVToAvro
- ConvertJSONToAvro
- ConvertJSONToSQL
- ConvertRecord
- CreateHadoopSequenceFile
- CryptographicHashAttribute
- CryptographicHashContent
- DeleteAzureBlobStorage
- DeleteAzureDataLakeStorage
- DeleteByQueryElasticsearch
- DeleteDynamoDB
- DeleteElasticsearch5
- DeleteGCSObject
- DeleteGridFS
- DeleteHBaseCells
- DeleteHBaseRow
- DeleteHDFS
- DeleteS3Object
- DeleteSQS
- DetectDuplicate
- DistributeLoad
- DuplicateFlowFile
- EncryptContent
- EnforceOrder
- EvaluateJsonPath
- EvaluateXPath
- EvaluateXQuery
- ExecuteGroovyScript
- ExecuteInfluxDBQuery
- ExecuteProcess
- GetHDFS
- GetHDFSFileInfo
- GetHDFSSequenceFile
- GetHTMLElement
- GetHTTP
- GetIgniteCache
- GetJMSQueue
- GetJMSTopic
- GetKafka
- GetMongoRecord
- GetSFTP
- GetSolr
- GetSplunk
- GetSQS
- GetTCP
- GetTwitter
- HandleHttpRequest
- HandleHttpResponse
- HashAttribute
- HashContent
- IdentifyMimeType
- InvokeAWSGatewayApi
- InvokeHTTP
- InvokeScriptedProcessor
- JoltTransformJSON
- JoltTransformRecord
- JsonQueryElasticsearch
- ListAzureBlobStorage
- ListDatabaseTables
- ListenHTTP
- ListenRELP
- ListenSyslog
- ListenTCP
- ListenTCPRecord
- ListenUDP
- ListenUDPRecord
- ListenWebSocket
- ListFile
- ListFTP
- ListGCSBucket
- ListHDFS
- ListS3
- ListSFTP
- LogAttribute
- LogMessage
- LookupAttribute
- LookupRecord
- MergeContent
- MergeRecord
- ModifyHTMLElement
- PutElasticsearchRecord
- PutEmail
- PutFile
- PutFTP
- PutGCSObject
- PutGridFS
- PutHBaseCell
- PutHBaseJSON
- PutHBaseRecord
- PutHDFS
- PutHive3QL
- PutHive3Streaming
- PutHiveQL
- PutHiveStreaming
- PutHTMLElement
- PutInfluxDB
- PutJMS
- PutKafka
- PutKinesisFirehose
- PutKinesisStream
- PutKudu
- PutLambda
- PutMongoRecord
- PutORC
- PutParquet
- PutRecord
- PutRiemann
- PutS3Object
- PutSFTP
- PutSNS
- PutSolrContentStream
- PutSolrRecord
- PutSplunk
- PutSQL
- PutSQS
- PutSyslog
- PutTCP
- PutUDP
- PutWebSocket
- QueryCassandra
- QueryDatabaseTable
- QueryDatabaseTableRecord
- QueryElasticsearchHttp
- QueryRecord
- QuerySolr
- QueryWhois
- ReplaceText
- ReplaceTextWithMapping
- ResizeImage
- RetryFlowFile

- ExecuteScript
- ExecuteSQL
- ExecuteSQLRecord
- ExecuteStreamCommand
- ExtractAvroMetadata
- ExtractGrok
- ExtractHL7Attributes
- ExtractImageMetadata
- ExtractText
- FetchAzureBlobStorage
- FetchDistributedMapCache
- FetchElasticsearch
- FetchElasticsearch5
- FetchElasticsearchHttp
- FetchFile
- FetchFTP
- FetchGCSObject
- FetchGridFS
- FetchHBaseRow
- FetchHDFS
- FetchParquet
- FetchS3Object
- FetchSFTP
- FlattenJson
- ForkRecord
- GenerateFlowFile
- GenerateTableFetch
- GeoEnrichIP
- MonitorActivity
- Notify
- ParseCEF
- ParseEvtx
- ParseSyslog
- PartitionRecord
- PostHTTP
- PublishAMQP
- PublishGCPubSub
- PublishJMS
- PublishKafka
- PublishKafka_0_10
- PublishKafka_1_0
- PublishKafka_2_0
- PublishKafkaRecord_0_10
- PublishKafkaRecord_1_0
- PublishKafkaRecord_2_0
- PublishMQTT
- PutAzureBlobStorage
- PutAzureDataLakeStorage
- PutAzureEventHub
- PutAzureQueueStorage
- PutBigQueryBatch
- PutBigQueryStreaming
- PutCassandraQL
- PutCloudWatchMetric
- PutCouchbaseKey
- PutDatabaseRecord
- RouteHL7
- RouteOnAttribute
- RouteOnContent
- RouteText
- ScanAttribute
- ScanContent
- ScanHBase
- ScrollElasticsearchHttp
- SegmentContent
- SelectHive3QL
- SelectHiveQL
- SplitAvro
- SplitContent
- SplitJson
- SplitRecord
- SplitText
- SplitXml
- TagS3Object
- TailFile
- TransformXml
- UnpackContent
- UpdateAttribute
- UpdateCounter
- UpdateRecord
- ValidateCsv
- ValidateRecord
- ValidateXml
- Wait
- YandexTranslate

Supported NiFi Controller Services

Apache NiFi 1.11.4 ships with 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.

- AvroReader
- AvroRecordSetWriter
- AvroSchemaRegistry
- AWSCredentialsProviderControllerService
- AzureStorageCredentialsControllerService
- AzureStorageCredentialsControllerService
- CassandraSessionProvider
- CouchbaseClusterService
- CouchbaseKeyValueLookupService
- ElasticSearchLookupService
- ElasticSearchStringLookupService
- FreeFormTextRecordSetWriter
- GCPCredentialsControllerService
- GrokReader
- HBaseLookup_2_ClientMapCacheService
- HBase_1_1_2_ClientService
- HBase_1_1_2_ListLookupService
- HBase_1_1_2_RecordLookupService
- ParquetReader
- ParquetRecordSetWriter
- PrometheusRecordSink
- RedisConnectionPoolService
- RedisDistributedMapCacheClientService
- RestLookupService
- ScriptedLookupService
- ScriptedReader
- ScriptedRecordSetWriter

- CouchbaseMapCacheClient
- CouchbaseRecordLookupService
- CSVReader
- CSVRecordLookupService
- CSVRecordSetWriter
- DatabaseRecordLookupService
- DatabaseRecordSink
- DBCPConnectionPool
- DBCPConnectionPoolLookup
- DistributedMapCacheClientService
- DistributedMapCacheLookupService
- DistributedMapCacheServer
- DistributedSetCacheClientService
- DistributedSetCacheServer
- DruidTranquilityController
- ElasticSearchClientServiceImpl
- HBase_2_ClientMapCacheService
- HBase_2_ClientService
- HBase_2_RecordLookupService
- Hive3ConnectionPool
- HiveConnectionPool
- HortonworksSchemaRegistry
- JMSConnectionFactoryProvider
- JndiJmsConnectionFactoryProvider
- JsonPathReader
- JsonRecordSetWriter
- JsonTreeReader
- KafkaRecordSink_1_0
- KafkaRecordSink_2_0
- KeytabCredentialsService
- KuduLookupService
- ScriptedRecordSink
- SimpleDatabaseLookupService
- SimpleKeyValueLookupService
- SimpleScriptedLookupService
- SiteToSiteReportingRecordSink
- StandardHttpContextMap
- StandardProxyConfigurationService
- StandardRestrictedSSLContextService
- StandardS3EncryptionService
- StandardSSLContextService
- Syslog5424Reader
- SyslogReader
- VolatileSchemaCache
- XMLReader
- XMLRecordSetWriter

Supported NiFi Reporting Tasks

Apache NiFi 1.11.4 ships with 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

Unsupported Features in Cloudera DataFlow for Data Hub 7.2.7

Some features exist within Cloudera DataFlow for Data Hub 7.2.7 components, but are not supported by Cloudera.

Unsupported Flow Management features

There are no unsupported Flow Management features in Cloudera DataFlow for Data Hub 7.2.7

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 7.2.7, but are not supported by Cloudera.

Kafka

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

- Only Java and .Net based clients are supported. Clients developed with C, C++, Python, and other languages are currently not supported.
- While Kafka Connect is available as part of Runtime, it is currently not supported in CDP Public Cloud. NiFi is a proven solution for batch and real time data loading that complement Kafka's message broker capability. For more information, see [Creating your first Flow Management cluster](#).
- The Kafka default authorizer is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.

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.

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 7.2.7, but are not supported by Cloudera.

Flink

The following Flink 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*.

- SQL Client

Related Information

[Cloudera Community Forum](#)

Apache Patch Information in Cloudera DataFlow for Data Hub 7.2.7

The following sections list patches in each Cloudera DataFlow in Data Hub component, beyond what was fixed in the base version of the Apache component.

NiFi patches

This release provides Apache NiFi 1.11.4 and these additional Apache patches.

- [NIFI-7422](#) and [NIFI-7452](#) - Improvements for Atlas lineage with AWS S3 and Azure ADLS
- [NIFI-7436](#) - Performance improvements on analytics predictions
- [NIFI-7683](#) and [NIFI-7994](#) - ReplaceText issue
- [NIFI-7707](#) - Google Cloud PubSub issue
- [NIFI-7968](#) and [NIFI-7954](#) - Kerberos relogin issues

NiFi Registry patches

This release provides Apache NiFi Registry 0.5.0. There are no additional Apache patches.

Known Issues In Cloudera DataFlow for Data Hub 7.2.7

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera DataFlow for Data Hub 7.2.7.

Known Issues in Flow Management

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

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](#)

Known Issues in Streams Messaging

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

Kafka

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

Known Issues

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

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

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

Certain Kafka command line tools require direct access to Zookeeper

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

- `kafka-reassign-partitions`

None

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

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

None

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

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

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

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

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

None

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

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 in Schema Registry in this release:

CDPD-21877Schema Registry serde jar upload fails on AWS and GCP (NoClassDefFoundError)

Uploading serializer JAR files for schemas does not work in AWS/Azure/GCP in CDP 7.2.7.

Streams Messaging Manager

There are no known issues in Streams Messaging Manager in this release.

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-14019: SRM may automatically re-create deleted topics

If `auto.create.topics.enable` is enabled, deleted topics are automatically recreated on source clusters.

Prior to deletion, remove the topic from the topic whitelist with the `srm-control` tool. This prevents topics from being re-created.

```
srm-control topics --source [SOURCE_CLUSTER] --target [TARGET_CLUSTER] --remove [TOPIC1][TOPIC2]
```

CDPD-13864 and CDPD-15327: Replication stops after the network configuration of a source or target cluster is changed

If the network configuration of a cluster which is taking part in a replication flow is changed, for example, port numbers are changed as a result of enabling or disabling TLS, SRM will not update its internal configuration even if SRM is reconfigured and restarted. From SRM's perspective, it is the cluster identity that has changed. SRM cannot determine whether the new identity corresponds to the same cluster or not, only the owner or administrator of that cluster can know. In this case, SRM tries to use the last known configuration of that cluster which might not be valid, resulting in the halt of replication.

The internal topic storing the configuration of SRM can be deleted. After a restart SRM will re-create and re-populate it with the configuration data loaded from its property file. The topic is hosted on the target cluster of the replication flow. The topic name is: `mm2-configs.[SOURCE_ALIAS].internal`. However, changing a replicated cluster's identity is generally not recommended.

CDPD-22094: The SRM service role displays as healthy, but no metrics are processed

The SRM service role might encounter errors that make metrics processing impossible. An example of this is when the target Kafka cluster is not reachable. The SRM service role does not automatically stop or recover if such an error is encountered. It continues to run and displays as healthy in Cloudera Manager. Metrics, however, are not processed. In addition, no new data is displayed in SMM for the replications.

1. Ensure that all clusters are available and are in a healthy state.
2. Restart SRM.

CDPD-22389: The SRM driver role displays as healthy, but replication fails

During startup, the SRM driver role might encounter errors that make data replication impossible. An example of this is when one of the clusters added for replication is not reachable. The SRM driver role does not automatically stop or recover if such an error is encountered. It will start up, continue to run, and display as healthy in Cloudera Manager. Replication, however, will not happen.

1. Ensure that all clusters are available and are in a healthy state.

2. Restart SRM.

CDPD-23683: The replication status reported by the SRM service role for healthy replications is flaky

The replication status reported by the SRM service role is flaky. The replication status might change between active and inactive frequently even if the replication is healthy. This status is also reflected in SMM on the replications tab.

None

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 `[***SOURCE CLUSTER ALIAS***]->[***TARGET CLUSTER ALIAS***].consumer.isolation.level=read_committed` to the Streams Replication Manager's Replication Configs SRM service property in Cloudera Manger.

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.

Known Issues in Streaming Analytics

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

There are no known Streaming Analytics issues in Cloudera DataFlow for Data Hub 7.2.7.

Fixed Issues in Cloudera DataFlow for Data Hub 7.2.7

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 7.2.7.

Fixed Issues in Flow Management

Review the list of Flow Management issues that are resolved in Cloudera DataFlow for Data Hub 7.2.7.

There are no fixed issues in this Flow Management release.

Fixed Issues in Streams Messaging

Review the list of Streams Messaging issues that are resolved in Cloudera DataFlow for Data Hub 7.2.7.

Kafka

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

This issue is now resolved.

OPSAPS-59032: Obsolete TLS protocols are enabled by default

TLS v1.0 and v1.1 connections to Kafka are no longer allowed by default.

Streams Messaging Manager

CDPD-20654: SMM fails to show replication flow status on the Replications tab

As a result of an SRM Service REST API issue, replication flow status as well as the latest cluster and topic level metrics were not provided to SMM. This caused SMM to show no data for throughput, replication latency, and checkpoint latency even if replication was working properly.

This issue is fixed. SMM now correctly displays this information.

CDPD-19045: Kerberos rewrite rules should be applied for SMM Ranger Authorization

Although SMM's Cloudera Manager config streams.messaging.manager.kerberos.name.rules is configured, this was not being honored when authorizing against Ranger.

CDPD-19495: SMM UI does not show producer data on topics page

The topics page, the topic profile pages, and the broker profile pages now show correct producer data.

Schema Registry

CDPD-18943: Remove HA tables from the database

Removed table "host_config" which was used for communication between hosts. Since the High-Availability feature was removed in 0.10.0, this table is not used anymore.

CDPD-17299: Schema Registry Client cannot relogin with dynamic JAAS config

When the Schema Registry Client used dynamic JAAS, it could not renew the Kerberos ticket after it expired. This issue has been fixed.

Streams Replication Manager

There are no fixed Streams Replication Manager issues in this release.

Fixed Issues in Streaming Analytics

Review the list of Streaming Analytics issues that are resolved in Cloudera DataFlow for Data Hub 7.2.7.

There are no fixed issues for Streaming Analytics in Cloudera DataFlow for Data Hub 7.2.7.