Cloudera Runtime 7.1.9

Kafka Connect Connector Reference

Date published: 2019-08-22
Date modified: 2023-09-07

CLOUDERA

Legal Notice

Contents

HTTP Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the HTTP Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Authorized Issuer DN Pattern

Description

A regular expression to apply against the Issuer's Distinguished Name of incoming connections. You can use this property when the Client Authentication property is set to REQUIRED (mutual TLS).

Default Value

.*

Accepted Values

Required

false

Authorized Subject DN Patter

Description

A regular expression to apply against the Subject's Distinguished Name of incoming connections. You can use this property when the Client Authentication property is set to REQUIRED (mutual TLS).

Default Value

.*

Accepted Values

Required

false

Base Path

Description

The base path or context path of the URL.

Default Value

contentListener

Accepted Values

Required

true

Client Authentication

Description

> The client authentication policy used for HTTPS.

Default Value

> NONE

Accepted Values

> NONE, REQUIRED

Required

> true

Keystore Filename

Description

> The fully-qualified filename of a keystore.
>
> This keystore is used to establish a secure connection between this connector and its clients using HTTPS.

Default Value

Accepted Values

Required

> true

Keystore Key Password

Description

> The password used to access the key stored in the keystore file configured in the Keystore Filename property.

Default Value

Accepted Values

Required

> true

Keystore Password

Description

> The password used to access the contents of the keystore configured in the Keystore Filename property.

Default Value

Accepted Values

Required

> true

Keystore Type

Description

> The type of the keystore configured in the Keystore Filename property.

Default Value

Accepted Values

> BCFKS, PKCS12, JKS

Required

>   true

Listening Port

Description

>   The port to listen on for communication.

Default Value

Accepted Values

Required

>   true

Truststore Filename

Description

>   The fully-qualified filename of a truststore.

>   When using one-way SSL (Client Authentication is set to NONE), this property must be completely
>   removed from the configuration JSON.

Default Value

Accepted Values

Required

>   false

Truststore Password

Description

>   The password used to access the contents of the truststore configured in the Truststore Filename
>   property.

>   When using one-way SSL (Client Authentication is set to NONE), this property must be completely
>   removed from the configuration JSON.

Default Value

Accepted Values

Required

>   false

Truststore Type

Description

>   The type of the truststore configured in the Truststore Filename property.

>   When using one-way SSL (Client Authentication is set to NONE), this property must be completely
>   removed from the configuration JSON.

Default Value

Accepted Values

>   BCFKS, PKCS12, JKS

Required

>   false

JDBC Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the JDBC Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Additional WHERE clause

Description

Specifies a custom condition to add in the WHERE clause of the SQL query.

Default Value

Accepted Values

Required

false

Columns to Return

Description

A comma-separated list of column names to be returned by the query. All columns are returned if it is not specified.

Default Value

Accepted Values

Required

false

Custom Query

Description

Specifies a custom SQL query to use instead of generating the query from the table and column names.

Default Value

Accepted Values

Required

false

Database Connection URL

Description

The database specific connection URL used for connecting to the database. For example, jdbc:pos tgresql://localhost:5432/postgres.

Default Value

Accepted Values

Required

> true

Database Driver Class Name

Description

> The database driver class name For example, org.postgresql.Driver.

Default Value

Accepted Values

Required

> true

Database Driver Location

Description

> A comma-separated list of files or folders containing the JDBC client libraries.

Default Value

Accepted Values

Required

> true

Database Table Name

Description

> The name of the database table to query.

Default Value

Accepted Values

Required

> true

Database Type

Description

> The database type used for generating database specific SQL queries.

Default Value

> Generic

Accepted Values

> Generic, Oracle, Oracle 12+, MS SQL 2008, MS SQL 2012+, MySQL, PostgreSQL

Required

> true

Database User Name

Description

> The database user name. If username/password authentication is not required by the database server, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Database User Password

Description

> The database user password. If username/password authentication is not required by the database
> server, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Date Format

Description

> Specifies the format to use when writing Date fields to JSON.

Default Value

> yyyy-MM-dd

Accepted Values

Required

> true

Initial Load Strategy

Description

> Specifies how existing rows in the database table are handled when the connector is started for the
> first time.

Default Value

> Start at Beginning

Accepted Values

> Start at Beginning, Start at Current Maximum Values

Required

> true

Kafka Message Data Format

Description

> Specifies the format of the messages the connector sends to Kafka. The database row is converted
> to this format. If the output format is Avro and Schema Access Strategy is set to Inherit Schema,
> the schema is embedded in the output message. If the output format is Avro and Schema Access
> Strategy is set to Schema Registry, the schema is not embedded in the output message.

Default Value

> Avro

Accepted Values

> Avro, JSON

Required

> true

Kafka Message Key Column

Description

Specifies a database table column. The value of the column specified is used as the key of the Kafka message.

Default Value

Accepted Values

Required

false

Kerberos Keytab

Description

The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

default

Accepted Values

Required

true

Maximum-value Columns

Description

A comma-separated list of column names used for incremental loading.

Default Value

Accepted Values

Required

false

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the database record.

- If set to Inherit Schema, the schema is determined from the database schema.
- If set to Schema Registry, the schema is read from Schema Registry.
- If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.

Default Value

Inherit Schema

Accepted Values

Inherit Schema, Schema Registry, HWX Content-Encoded Schema Reference

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry. If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Time Format

Description

> Specifies the format to use when writing Time fields to JSON.

Default Value

> HH:mm:ss

Accepted Values

Required

> true

Timestamp Format

Description

> Specifies the format to use when writing Timestamp fields to JSON.

Default Value

> yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename

Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection
> with Schema Registry using HTTPS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

> tue

Truststore Password

Description

> The password used to access the contents of the truststore configured in the Truststore Filename
> property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type

Description

> The type of the truststore configured in the Truststore Filename property.

Default Value

> JKS

Accepted Values

> BCFKS, PKCS12, JKS

Required

> true

JMS Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the JMS Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

JMS Broker URI
Description

> The URI to use to connect to the JMS Message Broker. For example, ssl://myhost:61617

Default Value

Accepted Values

Required

> true

JMS Client Libraries
Description

> The path to the directory containing the JMS client libraries.

Default Value

Accepted Values

Required

> true

JMS Connection Factory Class Name
Description

> The fully qualified name of the JMS ConnectionFactory implementation class. For example, org. apache.activemq.ActiveMQSslConnectionFactory.

Default Value

Accepted Values

Required

true

JMS Destination Name

Description

The name of the JMS Destination.

Default Value

Accepted Values

Required

true

JMS Destination Type

Description

The type of the JMS Destination.

Default Value

QUEUE

Accepted Values

QUEUE, TOPIC

Required

true

JMS User Name

Description

The username used for authentication.

If username/password authentication is not required by the JMS Message Broker, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

JMS User Password

Description

The password used for authentication.

If username/password authentication is not required by the JMS Message Broker, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Keystore Filename

Description

The fully-qualified filename of a keystore. This keystore is used for mutual TLS towards the JMS Message Broker.

If the JMS Message Broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Keystore Key Password

Description

The password used to access the key stored in the keystore file configured in the Keystore Filename property.

If the JMS Message Broker broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Keystore Password

Description

The password used to access the contents keystore configured in the Keystore Filename property.

If the JMS Message Broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Keystore Type

Description

The type of the keystore configured in the Keystore Filename property.

If the JMS Message Broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

BCFKS, PKCS12, JKS

Required

false

Truststore Filename

Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with the JMS server using TLS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Truststore Password
Description

The password used to access the contents of the truststore configured in the Truststore Filename property.

Default Value

password

Accepted Values

Required

true

Truststore Type
Description

The type of the truststore configured in the Truststore Filename property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

MQTT Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the MQTT Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Keystore Filename
Description

The fully-qualified filename of a keystore. This keystore is used for mutual TLS towards the MQTT server.

If the MQTT broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Keystore Key Password

Description

> The password used to access the key stored in the keystore file configured in the Keystore Filename property.
>
> If the MQTT broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Keystore Password

Description

> The password used to access the contents keystore configured in the Keystore   Filename property.
>
> If the MQTT broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Keystore Type

Description

> The type of the keystore configured in the Keystore Filename property.
>
> If the MQTT broker does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

> BCFKS, PKCS12, JKS

Required

> false

MQTT Broker URI

Description

> The URI to use to connect to the MQTT broker.
>
> Example URI if SSL is not used (keystore-related parameters are removed and and the default truststore is used): tcp://localhost:1883
>
> Example URI if SSL is used: ssl://localhost:8883

Default Value

> tcp://localhost:1883

Accepted Values

Required

> true

MQTT Password

Description

> Password to use when connecting to the MQTT broker.
>
> If username-password authentication is not required by the MQTT broker, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

MQTT Quality of Service

Description

> The Quality of Service (QoS) to receive the message with.
>
> 0 - at most once
>
> 1 - at least once
>
> 2 - exactly once

Default Value

> 0

Accepted Values

> 0,1 or 2

Required

> true

MQTT Topics

Description

> Specifies the MQTT topic to subscribe to. Use an MQTT wildcard to subscribe to multiple topics simultaneously.

Default Value

Accepted Values

Required

> true

MQTT Username

Description

> Username to use when connecting to the MQTT broker.
>
> If username-password authentication is not required by the MQTT broker, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

        false

Truststore Filename

Description

        The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with the MQTT server using TLS.

Default Value

        The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

        true

Truststore Password

Description

        The password used to access the contents of the truststore configured in the Truststore Filename property.

Default Value

        password

Accepted Values

Required

        true

Truststore Type

Description

        The type of the truststore configured in the Truststore Filename property.

Default Value

        JKS

Accepted Values

        BCFKS, PKCS12, JKS

Required

        true

SFTP Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the SFTP Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Avro Schema Write Strategy

Description

Specifies whether the record schema is written to the output Avro file.

Default Value

Embed Avro Schema

Accepted Values

Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

false

CSV Character Set

Description

The character set used to read the input CSV files.

This property is ignored if the input is not a CSV file or if record processing is not enabled.

Default Value

UTF-8

Accepted Values

Required

true

CSV Escape Character

Description

The escape character used in the input CSV files to escape other special characters.

This property is ignored if the input is not a CSV file or if record processing is not enabled.

Default Value

\

Accepted Values

Required

true

CSV Quote Character

Description

The quote character used in the input CSV files.

This property is ignored if the input is not a CSV file or if record processing is not enabled.

Default Value

"

Accepted Values

Required

true

CSV Record Separator

Description

The record separator used in the input CSV files.

This property is ignored if the input is not a CSV file or if record processing is not enabled.

Default Value

\n

Accepted Values

Required

true

CSV Treat First Line as Header

Description

Specifies whether the first line in the input file is handled as a header.

Ignored if the input is not a CSV file or record processing is not enabled.

Default Value

false

Accepted Values

true, false

Required

true

CSV Trim Fields

Description

Specifies whether whitespace characters are removed from the beginning and the end of fields.

Ignored if the input is not a CSV file or record processing is not enabled.

Default Value

true

Accepted Values

true, false

Required

true

CSV Value Separator

Description

The value separator used in the input CSV files.

Ignored if the input is not a CSV file or record processing is not enabled.

Default Value

,

Accepted Values

Required

true

Completion Strategy

Description

Specifies what to do with the original file on the server once it has been fetched.

If the Completion Strategy fails, a warning is logged but the data is still transferred.

Default Value

None

Accepted Values

None, Move File, Delete File

Required

true

Date Format

Description

Specifies the format used for parsing date fields in the input data.

This property is only used if Input Data Format is set to CSV or JSON.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Enable Record Processing

Description

Enables or disables record processing.

If set to true, the Input Data Format is considered and the file gets parsed into records. In this case the Records Per Kafka Message property defines how many records are written into one Kafka message.

If set to false, the entire file gets forwarded to Kafka as one message.

Default Value

true

Accepted Values

true, false

Required

true

File Filter Regex

Description

The Java regular expression to use for filtering filenames. Only files whose names match the regular expression are fetched.

Default Value

.*

Accepted Values

Required

true

Follow Symlink

Description

If set to true, both symbolic files and nested symbolic subdirectories are pulled. Otherwise, symbolic files are not read and symbolic link subdirectories are not traversed.

Default Value

false

Accepted Values

true, false

Required

true

Grok Expression

Description

Specifies the format of a line in Grok format. This allows the connector to understand how to parse each line in the input file. If a line in the file does not match this pattern, the line is handled according to what is set in the Grok No Match   Behavior property.

A valid Grok expression must be specified using this property even if Grok format is not used.

Default Value

%{GREEDYDATA:message}

Accepted Values

Required

true

Grok No Match Behavior

Description

Specifies how to handle lines that do not match the pattern set in the Grok   Expression property.

If set to append-to-previous-message, non-matching lines are appended to the last field of the previous message.

If set to skip-line, non-matching lines are skipped.

If set to raw-line, non-matching lines are only added to the _raw field.

Default Value

append-to-previous-message

Accepted Values

append-to-previous-message, skip-line, raw-line

Required

true

Host Key File

Description

The fully-qualified filename of the host key file.

If supplied, this file is used as the host key.

If a host key is not supplied, but Strict Host Key Checking is set to true, the known_hosts and known_hosts2 files from the ~/.ssh directory are used.

If a host key is not supplied and Strict Host Key Checking is set to false, no host key file is used.

This parameter must either contain the fully-qualified name of a file, or be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Hostname
Description

The hostname or IP address of the remote system.

Default Value

localhost

Accepted Values

Required

true

Ignored Dotted Files
Description

Specifies whether to ignore files whose names begin with a dot (".").

Default Value

true

Accepted Values

true, false

Required

true

Input Data Format
Description

The format in which the input file contains record-oriented data.

If Enable Record Processing is set to false, this setting is ignored.

Default Value

JSON

Accepted Values

JSON, CSV, GROK

Required

true

Kerberos Keytab for Schema Registry
Description

The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Kerberos Principal for Schema Registry

Description

> The Kerberos principal used for authenticating to Schema Registry.

Default Value

> default

Accepted Values

Required

> true

Move Destination Directory

Description

> The fully-qualified name of the directory on the remote server to move the original file to once it is ingested. This property is ignored unless the Completion   Strategy property is set to Move File. The specified directory must already exist on the remote system.

> This parameter must either contain the fully-qualified name of a directory, or be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Password

Description

> The password to use when connecting to the SFTP server.

> If the server does not require a password, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Path Filter Regex

Description

> The Java Regular Expression to use for filtering paths.

> If Search   Recursively is set to true, only subdirectories whose path matches the given regular expression are scanned.

> If Search   Recursively is set to false, this property is ignored.

Default Value

> .*

Accepted Values

Required

true

Port

Description

The port that the remote system is listening on for file transfers.

Default Value

22

Accepted Values

Required

true

Private Key File

Description

The fully-qualified filename of a private key file.

If no private key is used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Private Key Password

Description

The password used to access the private key.

If no private key is used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Record Per Kafka Message

Description

Specifies how many records to write into each Kafka message.

If Enable Record Processing is set to false, this setting is ignored.

Default Value

1

Accepted Values

Required

true

Remote Path

Description

The path on the remote system from which to pull files.

Default Value

.

Accepted Values

Required

       true

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the input records if the Enable Record Pr
ocessing property is set to true .

The value you set here depends on the input data format.

If set to Schema Registry, the schema is read from Schema Registry.

This setting works with all input data formats.

If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if
your input data format is either JSON or CSV.

If set to Field Names From Grok   Expression, the schema is determined using the field names in
the Grok Expression property. This setting can only be used if your input data format is GROK.

Default Value

       Schema Registry

Accepted Values

       Schema Registry, Infer Schema, Field Names From Grok Expression

Required

       true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch
and Schema Version cannot be specified at the same time. If one is specified, the other needs to be
removed from the configuration. If Schema Registry is not used, this property must be completely
removed from the configuration.

Default Value

Accepted Values

Required

       false

Schema Name

Description

The schema name to look up in Schema Registry.

If the Schema   Access Strategy property is set to Schema Registry, this property must contain a
valid schema name.

If Schema Registry is not used, this property must be completely removed from the configuration
JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server.

If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

true

Search Recursively

Description

Specifies whether to pull files from arbitrarily nested subdirectories. Subdirectories are not traversed if set to false.

Default Value

false

Accepted Values

true, false

Required

true

Strict Host Key Checking

Description

Specifies whether strict enforcement of host keys is applied.

Default Value

false

Accepted Values

true, false

Required

true

Time Format

Description

Specifies the format used for parsing time fields in the input data. This property is only used if Input Data Format is set to CSV or JSON.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format

Description

Specifies the format used for parsing timestamp fields in the input data. This property is only used if Input Data Format is set to CSV or JSON.

Default Value

yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

true

Truststore Filename for Schema Registry

Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using TLS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Truststore Password for Schema Registry

Description

The password used to access the contents of the truststore configured in the Truststore Filename for Schema Registry property.

Default Value

password

Accepted Values

Required

true

Truststore Type for Schema Registry

Description

The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

Accepted Values

BCFKS, PKCS12, JKS

Required

true

Username

Description

Username for connecting to the SFTP server.

Default Value

Accepted Values

Required

true

Stateless NiFi Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the Stateless NiFi Source connector.

In addition to the properties listed here, this connector also accepts the properties of the Kafka Connect framework. For a comprehensive list of these properties, see the Apache Kafka documentation.

dataflow.timeout

Description

Specifies the maximum amount of time to wait for the dataflow to complete. If the dataflow does not complete before this timeout, the thread is interrupted and the dataflow is considered as a failure. The session is rolled back and the connector retriggers the flow. Defaults to 60 seconds if not specified.

Default Value

60 seconds

Accepted Values

Required

false

extensions.directory

Description

Specifies the directory that stores downloaded extensions. Extensions are the NAR (NiFi Archive) files containing the processors and controller services a flow might use. Since Stateless NiFi is only the NiFi engine, it does not contain any of the processors and controller services you might use in your flow. When deploying the connector with the custom flow, the system needs to download the specific extensions that your flow uses from Nexus (unless they are already present in this directory). These extensions are stored in this directory. Because the default directory might not be writable, and to aid in upgrade scenarios, Cloudera recommends that you always specify an extensions directory.

Default Value

/tmp/nifi-stateless-extensions

Accepted Values

Required

true

flow.snapshot

Description

Specifies the dataflow to run. When using SMM to deploy a connector, the value you set in this property must be a JSON object. URLs, file paths, or escaped JSON strings are not supported when using SMM. Alternatively, if using the Kafka Connect REST API to deploy a connector, this can be a file containing the dataflow, a URL that points to a dataflow, or a string containing the entire dataflow as an escaped JSON. Cloudera however, does not recommend using the Kafka Connect REST API to interact with this connector or Kafka Connect.

Default Value

Accepted Values

Required

true

header.attribute.regex

Description

A Java regular expression that is evaluated against all flowfile attribute names. Any attribute name matching the regular expression is converted into a Kafka message header. The name of the attribute is used as the header key, the value of the attribute is used as the header value. If not specified, headers are not added to the Kafka record.

Default Value

Accepted Values

Required

false

header.name.regex

Description

A Java regular expression that will be evaluated against all flowfile attribute names. For any attribute whose name matches the regular expression, the Kafka record will have a header whose name matches the attribute name and whose value matches the attribute value. If not specified, the Kafka record will have no headers added to it.

Default Value

Accepted Values

Required

false

key.attribute

Description

Specifies the name of a flowfile attribute that should be used to specify the key of the Kafka record. If not specified, the Kafka record will not have a key associated with it. If specified, but the attribute does not exist on a particular flowfile, it will also have no key associated with it.

Default Value

Accepted Values

Required

false

krb5.file

Description

Specifies the krb5.conf file to use if the dataflow interacts with any services that are secured using Kerberos. Defaults to /etc/krb5.conf if not specified.

Default Value

/etc/krb5.conf

Accepted Values

Required

false

name

Description

The name of the connector. On the SMM UI, the connector names are specified using the Enter Name field. The name that you enter in the Enter Name field is automatically set as the value of the name property when the connector is deployed. Because of this, the name property is omitted from the configuration template provided in SMM. If you manually add the name property to the configuration in SMM, ensure that the value you set matches the connector name specified in the Enter Name field. Otherwise, the connector fails to deploy.

Default Value

Accepted Values

Required

True

nexus.url

Description

Specifies the Base URL of the Nexus instance to source extensions from. If configuring a Nexus instance that has multiple repositories, include the name of the repository in the URL. For example, https://nexus-private.myorganization.org/nexus/repository/my-repository/. If the property is not specified, the necessary extensions (the ones used by the flow) must be provided in the extensions directory before deploying the connector.

Default Value

Accepted Values

Required

true

output.port

Description

The name of the output port in the NiFi dataflow to pull data from. If the dataflow contains exactly one port, this property is optional and can be omitted. However, if the dataflow contains multiple ports (for example, a success and a failure port), this property must be specified. If any flowfile is sent to any port other than the specified port, it is considered as a failure. The session is rolled back and no data is collected.

Default Value

Accepted Values

Required

false

parameter.[***FLOW PARAMETER NAME***]

Description

Specifies a parameter to use in the dataflow. For example, assume that you have the following entry in your connector configuration "parameter.Directory":   "/mydir". In a case like this, any parameter context in the dataflow that has a parameter named Directory gets the specified value (/mydir). If the dataflow has child process groups, and those child process groups have their own parameter contexts, the value is used for all parameter contexts that contain a parameter named Directory. Parameters can also be applied to specific Parameter Contexts only. This can be done by prefixing the parameter name (Directory) with the name of the parameter context followed by a colon. For example, parameter.My Context:Directory only applies the specified value for the Dire ctory parameter in the parameter context named My Context.

Default Value

Accepted Values

Required

false

topic.name.attribute

Description

Specifies the name of a flowfile attribute to use for determining which Kafka topic a flowfile is sent to. Either the topics or topic.name.attribute property must be specified. If both are specified, topi c.name.attribute takes precedence. However, if a flowfile does not have the specified attribute name, then the connector falls back to using the topics property.

Default Value

Accepted Values

Required

false

topics

Description

The name of the topic to deliver data to. All flowfiles are delivered to the topic specified here. However, it is also possible to determine the topic individually for each flowfile. To do this, ensure that the dataflow specifies the topic name in an attribute, and then use topic.name.attribute to specify the name of the attribute instead of topic name. For example, if you wanted a separate Kafka topic for each data source, you can omit the topics property and instead specify the attribute (for example, datasource.hostname) corresponding to the topic using the topic.name.attribute property.

Default Value

Accepted Values

Required

true

working.directory

Description

Specifies a directory on the Connect server that NiFi should use for unpacking extensions that it needs to perform the dataflow. The contents of extensions.directory are unpacked here. Defaults to / tmp/nifi-stateless-working if not specified.

Default Value

/tmp/nifi-stateless-working

Accepted Values

Required

false

Syslog TCP Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the Syslog TCP Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Authorized Issuer DN Pattern

Description

A regular expression that can be applied against the Issuer's Distinguished Name of incoming TLS connections.

Default Value

.*

Accepted Values

Required

false

Authorized Subject DN Pattern

Description

A regular expression that can be applied against the Subject's Distinguished Name of incoming TLS connections.

Default Value

.*

Accepted Values

Required

false

Avro Schema Write Strategy

Description

Specifies how the schema is attached to the outgoing Avro messages. This property only takes effect if the Output Format is AVRO.

- If set to Embed Avro Schema then the schema is embedded in every output Avro message.

- If set to Do Not Write Schema then no schema information is attached to the output Avro messages.
- If set to HWX Content-Encoded Schema Reference then a reference to the schema (identified by Schema Name) within Schema Registry is encoded in the content of the outgoing Avro messages.

Default Value

Embed Avro Schema

Accepted Values

Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

false

Character Set

Description

The character set used in the input as well as the output data.

Default Value

UTF-8

Accepted Values

Required

true

Client Authentication

Description

The client authentication policy used for SSL.

Default Value

REQUIRED

Accepted Values

NONE, WANT, REQUIRED

Required

true

Date Format

Description

Specifies the format used for writing date fields if the Output   Format is JSON. Otherwise this parameter is not used.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Grok Expression

Description

Specifies the format of a line in Grok format. This allows the connector to understand how to parse each line in the input file. If a line in the file does not match this pattern, the line is handled

according to the Grok No Match Behavior. A valid Grok expression must be specified using this property even if Grok format is not used.

Default Value

%{GREEDYDATA:message}

Accepted Values

Required

true

Grok No Match Behaviour

Description

Specifies how to handle lines that do not match the pattern set in the Grok Expression property.

- If set to append-to-previous-message, non-matching lines are appended to the last field of the previous message.
- If set to skip-line, non-matching lines are skipped. If set to raw-line, non-matching lines are only added to the _raw field.

Default Value

append-to-previous-message

Accepted Values

append-to-previous-message, skip-line, raw-line

Required

true

Input Data Format

Description

The format of incoming messages.

Default Value

Syslog 3164

Accepted Values

Syslog 3164, Syslog 5424, Grok

Required

true

Kerberos Keytab for Schema Registry

Description

The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal for Schema Registry

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

> default

Accepted Values

Required

> true

Max Batch Size

Description

> The maximum number of messages to add to a single batch of messages. If multiple messages are available, they are concatenated with a new line character up to this configured maximum number of messages.

Default Value

> 1

Accepted Values

Required

> true

Max Number of Worker Threads

Description

> he maximum number of worker threads available for handling TCP connections.

Default Value

> 2

Accepted Values

Required

> true

Output Format

Description

> The format of the messages written to Kafka.

Default Value

> AVRO

Accepted Values

> TEXT, AVRO, JSON

Required

> true

Output Grouping for JSON

Description

> Specifies how JSON objects are grouped in the connector output.
>
> - If set to output-array, the output will consist of an array of JSON objects.
> - If set to output-oneline, each line of the output data will be one JSON object. That is, each JSON object occupies one line in the output.

Default Value

> output-oneline

Accepted Values

output-array, output-oneline

Required

true

Port

Description

The port to listen on for communication.

Type

int

Default Value

Accepted Values

Required

true

SSL Keystore Filename

Description

The fully-qualified filename of a keystore. This keystore is used to establish a secure connection between this connector and its clients using SSL.

Default Value

Accepted Values

Required

true

SSL Keystore Key Password

Description

The password used to access the key stored in the keystore file configured in SSL Keystore Fil ename.

Default Value

Accepted Values

Required

true

SSL Keystore Password

Description

The password used to access the contents of the keystore configured in the SSL   Keystore Filen ame property.

Default Value

Accepted Values

Required

true

SSL Keystore Type

Description

The type of the keystore configured in the SSL Keystore Filename property.

Default Value

Accepted Values

>   BCFKS, PKCS12, JKS

Required

>   true

SSL Truststore Filename

Description

>   The fully-qualified filename of a truststore. It can be used for establishing connections using mutual
>   TLS. When using one-way SSL (Client   Authentication parameter set to NONE), this parameter
>   must be removed completely from the configuration JSON.

Default Value

Accepted Values

Required

>   false

SSL Truststore Password

Description

>   The password used to access the contents of the truststore configured in the SSL Truststore Filen
>   ame property. When using one-way SSL (Client Authentication parameter set to NONE), this
>   parameter must be removed completely from the configuration JSON.

Default Value

Accepted Values

Required

>   false

SSL Truststore Type

Description

>   The type of the truststore configured in the SSL Truststore Filename property. When using one-way
>   SSL (Client Authentication parameter set to NONE), this parameter must be removed completely
>   from the configuration JSON.

Default Value

Accepted Values

>   BCFKS, PKCS12, JKS

Required

>   false

Schema Access Strategy

Description

>   Specifies the strategy used for determining the schema of the input message. This property only
>   takes effect if the input data format is GROK.
>
>   - If set to Schema Registry then the schema is read from Schema Registry.
>   - If set to Field Names from Grok Expression then the schema is determined using the field names
>   in the Grok Expression property.

Default Value

Schema Registry

Accepted Values

Schema Registry, Field Names From Grok Expression

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry.

- If the Schema Access Strategy property is set to Schema   Registry, this property must contain a valid schema name.
- If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Time Format
Description

> Specifies the format used for writing time fields if the Output   Format is JSON. Otherwise this parameter is not used.

Default Value

> HH:mm:ss

Accepted Values

Required

> true

Timestamp Format
Description

> Specifies the format used for writing timestamp fields if the Output   Format is JSON. Otherwise this parameter is not used.

Default Value

> yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename for Schema Registry
Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using TLS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Truststore Password for Schema Registry
Description

> The password used to access the contents of the truststore configured in the Truststore Filename for Schema Registry property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type for Schema Registry

Description

The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

Syslog UDP Source properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the Syslog UDP Source connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Source connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Source properties reference.

Avro Schema Write Strategy

Description

Specifies how the schema is attached to the outgoing Avro messages. This property only takes effect if the Output Format is AVRO.

- If set to Embed Avro Schema then the schema is embedded in every output Avro message.
- If set to Do Not Write Schema then no schema information is attached to the output Avro messages.
- If set to HWX Content-Encoded Schema Reference then a reference to the schema (identified by Schema Name) within Schema Registry is encoded in the content of the outgoing Avro messages.

Default Value

Embed Avro Schema

Accepted Values

Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

false

Character Set

Description

The character set used in the input as well as the output data.

Default Value

UTF-8

Accepted Values

Required

> true

Date Format

Description

> Specifies the format used for writing date fields if the Output   Format is JSON. Otherwise this
> parameter is not used.

Default Value

> yyyy-MM-dd

Accepted Values

Required

> true

Grok Expression

Description

> Specifies the format of a line in Grok format. This allows the connector to understand how to
> parse each line in the input file. If a line in the file does not match this pattern, the line is handled
> according to what is set in the Grok No Match Behavior property. A valid Grok expression must be
> specified using this property even if Grok format is not used.

Default Value

> %{GREEDYDATA:message}

Accepted Values

Required

> true

Grok No Match Behaviour

Description

> Specifies how to handle lines that do not match the pattern set in the Grok Expression property.
>
> - If set to append-to-previous-message, non-matching lines are appended to the last field of the
>   previous message.
> - If set to skip-line, non-matching lines are skipped. If set to raw-line, non-matching lines are only
>   added to the _raw field.

Default Value

> append-to-previous-message

Accepted Values

> append-to-previous-message, skip-line, raw-line

Required

> true

Input Data Format

Description

> The format of incoming messages.

Default Value

> Syslog 3164

Accepted Values

Syslog 3164, Syslog 5424, Grok

Required

true

Kerberos Keytab for Schema Registry

Description

The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal for Schema Registry

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

default

Accepted Values

Required

true

Output Format

Description

The format of the messages written to Kafka.

Default Value

AVRO

Accepted Values

TEXT, AVRO, JSON

Required

true

Output Grouping for JSON

Description

Specifies how JSON objects are grouped in the connector output.

- If set to output-array, the output will consist of an array of JSON objects.
- If set to output-oneline, each line of the output data becomes one JSON object. That is, each JSON object occupies one line in the output.

Default Value

output-oneline

Accepted Values

output-array, output-oneline

Required

true

Port
Description

The port to listen on for communication.

Default Value

Accepted Values

Required

true

Schema Access Strategy
Description

Specifies the strategy used for determining the schema of the input message. This property only takes effect if the input data format is GROK.

- If set to Schema Registry then the schema is read from Schema Registry.
- If set to Field Names from Grok Expression then the schema is determined using the field names in the Grok Expression property.

Default Value

Schema Registry

Accepted Values

Schema Registry, Field Names From Grok Expression

Required

true

Schema Branch
Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name
Description

The schema name to look up in Schema Registry.

- If the Schema Access Strategy property is set to Schema   Registry, this property must contain a valid schema name.
- If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL
Description

The URL of the Schema Registry server. If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version
Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Time Format
Description

Specifies the format used for writing time fields if the Output   Format is JSON. Otherwise this parameter is not used.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format
Description

Specifies the format used for writing timestamp fields if the Output   Format is JSON. Otherwise this parameter is not used.

Default Value

yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

true

Truststore Filename for Schema Registry
Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection
with Schema Registry using TLS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Truststore Password for Schema Registry

Description

The password used to access the contents of the truststore configured in the Truststore Filename for
Schema Registry property.

Default Value

password

Accepted Values

Required

true

Truststore Type for Schema Registry

Description

The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

ADLS Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the ADLS
Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect
framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the
SMM UI, all valid properties are presented in the default configuration template. You can view the configuration
template to get a full list of valid properties. In addition, for more information regarding the accepted properties not
listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

ADLS Account Key
Description

The Storage Account Key to use for authentication to Azure. Required when Account Key authentication is used.

Default Value

Accepted Values

Required

false

ADLS Account Name

Description

The name of the Azure Storage Account to use for authentication to Azure and the target account where the output file is uploaded.

Default Value

Accepted Values

Required

true

ADLS Directory Name

Description

The name of the directory where the output file is uploaded. Use an empty string to specify the root directory.

Default Value

Accepted Values

Required

true

ADLS Endpoint Suffix

Description

Specifies the endpoint suffix used for accessing the ADLS service.

Default Value

dfs.core.windows.net

Accepted Values

Required

true

ADLS Filesystem Name

Description

The name of the filesystem (or container) in the Azure Storage Account where the output file is uploaded.

Default Value

Accepted Values

Required

true

ADLS SAS Token

Description

> The SAS Token to use for authentication to Azure. Required when SAS Token authentication is used.

Default Value

Accepted Values

Required

> false

Avro Schema Write Strategy

Description

> Specifies how the record schema is attached to the output data file. Applicable only for Avro output (Output File Data Format is set to Avro).
>
> **Do Not Write Schema**
>
> Neither the schema nor reference to the schema is attached to the output Avro messages.
>
> **Embed Avro Schema**
>
> The schema is embedded in every output Avro message.
>
> **HWX Content-Encoded Schema Reference**
>
> A reference to the schema (identified by Schema Name) within Schema Registry is encoded in the content of the outgoing Avro messages.

Default Value

> Embed Avro Schema

Accepted Values

> Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

> false

Azure Service Principal Client ID

Description

> The Client ID of the Service Principal to use for authentication to Azure. Required when Service Principal authentication is used.

Default Value

Accepted Values

Required

> false

Azure Service Principal Client Secret

Description

> The Client Secret of the Service Principal to use for authentication to Azure. Required when Service Principal authentication is used.

Default Value

Accepted Values

Required

> false

Azure Service Principal Tenant ID

Description

The Tenant ID of the Service Principal to use for authentication to Azure. Required when Service Principal authentication is used.

Default Value

Accepted Values

Required

false

Date Format

Description

Specifies the format to use when writing date fields to JSON or CSV.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Kafka Message Data Format

Description

Specifies the format of the messages the connector receives from Kafka. If set to Avro or JSON, record processing is enabled. Raw can be used for unstructured text or binary data.

Default Value

Avro

Accepted Values

Avro, JSON, Raw

Required

true

Kerberos Keytab

Description

The fully-qualified filename of the Kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

default

Accepted Values

Required

> true

Maximum File Size

Description

> The maximum size of the output data file. No size limit is applied if this property is not specified. Example values: 100 MB, 1 GB.

Default Value

Accepted Values

Required

> false

Maximum Number of Entries

Description

> The maximum number of entries in the output data file. In the context of this property, entry can mean one of two things. If record processing is enabled (Output File Data Format is set to Avro or JSON), an entry is a record. Otherwise, entry means a Kafka message. Set this property to 1 if you want to create a separate output file for each Kafka message.

Default Value

> 1000000

Accepted Values

Required

> true

Output File Data Format

Description

> Specifies the format of the records written to the output file. Required when record processing is enabled (Kafka Message Data Format is set to Avro or JSON).

Default Value

> Avro

Accepted Values

> Avro, JSON, CSV

Required

> false

Output File Demarcator

Description

> Specifies the character sequence for demarcating (delimiting) message boundaries when multiple Kafka messages are ingested into an output file as raw messages (no record processing). This property can only be used if Kafka Message Data   Format is set to Raw. If you want to use newline as the delimiter, set this property to \n.

Default Value

Accepted Values

Required

> false

Output Filename Pattern

Description

Specifies the structure of the name of the output file. The pattern can contain string literal (fixed text) parts and one or more of the following expressions:

- ${filename.uuid}: Inserts a generated UUID in the filename.
- ${filename.timestamp}: Inserts the current timestamp in the filename.
- ${filename.sequence}: Inserts an incrementing sequence value in the filename.

In order to generate unique filenames, either ${filename.uuid} or ${filename.sequence} must be used in the pattern.

Examples:

- data_${filename.uuid}.json
- records_${filename.timestamp}_${filename.sequence}.avro

Default Value

${filename.uuid}

Accepted Values

Required

false

Output Filename Sequence Initial Value

Description

This property is used to configure the initial value of the ${filename.sequence} expression. The value you set in this property is not the initial value of the sequence. The sequence starts at the value of this property +1. For example, if you set this property to 0, the sequence starts at 1.

Default Value

0

Accepted Values

Required

false

Output Filename Sequence Padding Length

Description

Specifies the length of the ${filename.sequence} expression in characters. If the sequence has fewer characters than the value set in this property, it is padded with zeros (0). Padding is added to the left of the sequence.

Default Value

6

Accepted Values

Required

false

Output Filename Timestamp Format

Description

The timestamp format to use for the ${filename.timestamp} expression. For example, yyyyMMdd _HHmmss_SSS.

Default Value

Accepted Values

Required

> false

Schema Access Strategy

Description

> Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.
>
> - If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
> - If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
> - If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
> - If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.
>
> This property is not used if record processing is disabled (Kafka Message   Data Format is set to Raw).

Default Value

> Schema Registry

Accepted Values

> Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

> true

Schema Branch

Description

> The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Schema Name

Description

> The schema name to look up in Schema Registry. If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Time Format

Description

Specifies the format to use when writing Time fields to JSON or CSV.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format

Description

Specifies the format to use when writing Timestamp fields to JSON or CSV.

Default Value

yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

true

Truststore Filename

Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using HTTPS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Truststore Password

Description

> The password used to access the contents of the truststore configured in the Truststore Filename property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type

Description

> The type of the truststore configured in the Truststore Filename property.

Default Value

> JKS

Accepted Values

> BCFKS, PKCS12, JKS

Required

> true

Use Azure Managed Identity

Description

> Turns on authentication using Managed Identity.

Default Value

> false

Accepted Values

> true, false

Required

> true

Amazon S3 Sink properties reference

Amazon S3 Sink connector properties reference.

The following table collects connector properties that are specific for the Amazon S3 Sink Connector. For properties common to all sink connectors, see the upstream Apache Kafka documentation.

Property Name	Description	Type	Default Value	Accepted Values	Recommended Value
aws.s3.bucket	The target S3 bucket name.	String	none	Any valid S3 bucket name.	
aws.s3.service_endpoint	The target S3 host and port.	String	none	Any valid S3 endpoint.	
aws.access_key_id	The AWS secret key ID to authenticate.	String	none	Any valid secret key issued by AWS.	
aws.secret_access_key	The AWS secret access key to authenticate.	String	none	Any valid access key issued by AWS.	
value.converter	Value conversion class.	String	none		com.cloudera.dim.kafka.connect.converts
value.converter.passthrough.enabled	Configures whether the AvroConverter translates an Avro record into Kafka Connect Data or transparently passes the Avro encoded bytes as payload.	Boolean	true	true, false	True if input and output are both Avro.
value.converter.schema.registry.url	This is URL to the Schema Registry server.	String	none		
output.storage	The S3 storage implementation class.	String	none		com.cloudera.dim.kafka.connect.s3.S3Pa
output.writer	The output file writer which determines the type of file to be written. The value of this property should be the FQCN of a class that implements the PartitionWriter interface.	String	none	• com.cloudera.dim.kafka.connect.partition.writers.avro.AvroP • com.cloudera.dim.kafka.connect.partition.writers.json.JsonPa • com.cloudera.dim.kafka.connect.hdfs.parquet.ParquetPartition • com.cloudera.dim.kafka.connect.partition.writers.txt.TxtParti	com.cloudera.dim.kafka.connect.partition
output.avro.passthrough.enabled	Configures Whether the output writer expects an Avro encoded Kafka Connect data record. Must match the configuration of value.converter.passthrough.enabled.	Boolean	none	true, false	True if input and output are both Avro.

HDFS Sink properties reference

HDFS Sink connector properties reference.

The following table collects connector properties that are specific for the HDFS Sink Connector. For properties common to all sink connectors, see the upstream Apache Kafka documentation.

Property Name	Description	Type	Default Value	Accepted Values	Recommended Values
hdfs.uri	The file system URI to connect to on the destination cluster. This property supports any valid Hadoop-compatible filesystem (HCFS, For example, HDFS or ofs) URI.	String	None		
hdfs.output	The root directory on the HDFS cluster where all the output files will reside. The sub path has the following pattern: {topic}/{topic}_{par tition}_{endoffset}. {file extension}	String	/tmp		Any path on the HDFS file system where the role has read write permission.
hdfs.kerberos.authen tication	Enables or disables secure access to the HDFS cluster by authenticating with Kerberos.	Boolean	false	true or false	
hdfs.kerberos.user.p rincipal	The kerberos user principal.	String	null	The host-dependent Kerberos principal assigned to the Kafka Connect role.	
hdfs.kerberos.keytab .path	The path to the Kerberos keytab file.	String	null		In a Cloudera Manager provisioned environment, it's recommended to use the Cloudera Manager Config Provider to automatically provision the path.
hdfs.kerberos.nameno de.principal	The kerberos name node principal. Required when the HDFS cluster has data encryption on.	String	null		
hadoop.conf.path	The path to the site specific Hadoop configuration XML files. Required when the HDFS cluster has data encryption on.	String	null		
output.writer	The output file writer which determines the type of file to be written to the HDFS cluster. The value of this property should be the FQCN of a class that implements the PartitionWriter interface.	String	null	• com.cloudera.dim.kafka.connect.partition.writers.avro.AvroP • com.cloudera.dim.kafka.connect.partition.writers.json.JsonPa • com.cloudera.dim.kafka.connect.hdfs.parquet.ParquetPartition • com.cloudera.dim.kafka.connect.partition.writers.txt.TxtParti	com.cloudera.dim.kafka.connect.partition

Property Name	Description	Type	Default Value	Accepted Values	Recommended Values
output.avro.passthrough.enabled	Configures whether the output writer expects an Avro encoded Kafka Connect data record. Must match the configuration of value.converter.passthrough.enabled.	Boolean	true	true or false	True if input and output are both Avro.
value.converter	The converter to be used to translate the value field of the source Kafka record into Kafka Connect Data format.	String	Inherited from Kafka Connect worker properties.	• org.apache.kafka.connect.json.JsonConverter • org.apache.kafka.connect.storage.StringConverter • com.cloudera.dim.kafka.connect.converts.AvroConverter	com.cloudera.dim.kafka.connect.converts
value.converter.schema.registry.url	The URL to the Schema Registry server.	String	null	true or false	
value.converter.passthrough.enabled	Configures whether the AvroConverter translates an Avro record into Kafka Connect Data or transparently passes the Avro encoded bytes as payload.	Boolean	true	true or false	True if input and output are both Avro.

HDFS Stateless Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the HDFS Stateless Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Avro Schema Write Strategy

Description

Specifies how the record schema is attached to the output data file. Applicable only for Avro output (Output File Data Format is set to Avro).

Do Not Write Schema

Neither the schema nor reference to the schema is attached to the output Avro messages.

Embed Avro Schema

The schema is embedded in every output Avro message.

HWX Content-Encoded Schema Reference

A reference to the schema (identified by Schema Name) within Schema Registry is encoded in the content of the outgoing Avro messages.

Default Value

Embed Avro Schema

Accepted Values

Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

false

Compression Codec

Description

The codec used for file compression in HDFS. Use this property to set the codec if the output file format is JSON, Avro, or CSV.

Default Value

NONE

Accepted Values

NONE, DEFAULT, BZIP, GZIP, LZ4, LZO, SNAPPY, AUTOMATIC

Required

true

Compression Codec for Parquet

Description

Codec used for file compression in HDFS. Use this property to set the codec if the output file format is Parquet.

Default Value

UNCOMPRESSED

Accepted Values

UNCOMPRESSED,SNAPPY, GZIP, LZO

Required

true

Date Format

Description

Specifies the format to use when writing date fields to JSON or CSV.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Hadoop Configuration Resources

Description

A comma separated list of files which contains the Hadoop file system configuration.

Default Value

/etc/hadoop/conf/core-site.xml, /etc/hadoop/conf/hdfs-site.xml

Accepted Values

Required

> false

Kafka Message Data Format

Description

> Specifies the format of the messages the connector receives from Kafka. If set to Avro or JSON, record processing is enabled. Raw can be used for unstructured text or binary data.

Default Value

> Avro

Accepted Values

> Avro, JSON, Raw

Required

> true

Kerberos Keytab for HDFS

Description

> The fully-qualified filename of the Kerberos keytab associated with the principal for accessing HDFS.

Default Value

> The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Kerberos Keytab for Schema Registry

Description

> The fully-qualified filename of the Kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

> The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Kerberos Principal for HDFS

Description

> The Kerberos principal used for authenticating to HDFS.

Default Value

> default

Accepted Values

Required

> true

Kerberos Principal for Schema Registry

Description

> The Kerberos principal used for authenticating to Schema Registry.

Default Value

> default

Accepted Values

Required

> true

Maximum File Size

Description

> The maximum size of the output data file. No size limit is applied if this property is not specified. Example values: 100 MB, 1 GB.

Default Value

Accepted Values

Required

> false

Maximum Number of Entries

Description

> The maximum number of entries in the output data file. In the context of this property, entry can mean one of two things. If record processing is enabled (Output File Data Format is set to Avro or JSON), an entry is a record. Otherwise, entry means a Kafka message. Set this property to 1 if you want to create a separate output file for each Kafka message.

Default Value

> 1000000

Accepted Values

Required

> true

Output Directory Pattern

Description

> Specifies the full path of output HDFS directory. The pattern can contain string literals (fixed text) as well as the ${directory.timestamp} expression, which inserts the current timestamp in the directory name.

Default Value

Accepted Values

Required

> true

Output Directory Timestamp Format

Description

> The timestamp format to use for the ${directory.timestamp} expression. For example: yyyyMMdd.

Default Value

Accepted Values

Required

>   false

Output File Data Format

Description

>   Specifies the format of the records written to the output file. Required when record processing is
>   enabled (Kafka Message Data Format is set to Avro or JSON).

Default Value

>   Avro

Accepted Values

>   Avro, JSON, CSV, Parquet

Required

>   false

Output File Demarcator

Description

>   Specifies the character sequence for demarcating (delimiting) message boundaries when multiple
>   Kafka messages are ingested into an output file as raw messages (no record processing). This
>   property can only be used if Kafka Message Data   Format is set to Raw. If you want to use
>   newline as the delimiter, set this property to \n.

Default Value

Accepted Values

Required

>   false

Output Filename Pattern

Description

>   Specifies the structure of the name of the output file. The pattern can contain string literal (fixed
>   text) parts and one or more of the following expressions:

>   • ${filename.uuid}: Inserts a generated UUID in the filename.
>   • ${filename.timestamp}: Inserts the current timestamp in the filename.
>   • ${filename.sequence}: Inserts an incrementing sequence value in the filename.

>   In order to generate unique filenames, either ${filename.uuid} or ${filename.sequence} must be
>   used in the pattern.

>   Examples:

>   • data_${filename.uuid}.json
>   • records_${filename.timestamp}_${filename.sequence}.avro

Default Value

> ${filename.uuid}

Accepted Values

Required

>   false

Output Filename Sequence Initial Value

Description

This property is used to configure the initial value of the ${filename.sequence} expression. The value you set in this property is not the initial value of the sequence. The sequence starts at the value of this property +1. For example, if you set this property to 0, the sequence starts at 1.

Default Value

0

Accepted Values

Required

false

Output Filename Sequence Padding Length

Description

Specifies the length of the ${filename.sequence} expression in characters. If the sequence has fewer characters than the value set in this property, it is padded with zeros (0). Padding is added to the left of the sequence.

Default Value

6

Accepted Values

Required

false

Output Filename Timestamp Format

Description

The timestamp format to use for the ${filename.timestamp} expression. For example, yyyyMMdd _HHmmss_SSS.

Default Value

Accepted Values

Required

false

Output Files by Kafka Partitions

Description

Controls how fetched messages are merged together in the output file. If set to true, only messages fetched from the same Kafka partition get merged together in one output file (provided that Maxi mum Number of Entries is greater than 1). If set to false, messages are merged together in the order they are fetched from the Kafka topic. This property has no effect if Maximum Number of Entries is 1.

Default Value

false

Accepted Values

true, false

Required

false

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.

- If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
- If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
- If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
- If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.

This property is not used if record processing is disabled (Kafka Message   Data Format is set to Raw).

Default Value

Schema Registry

Accepted Values

Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry. If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Time Format

Description

Specifies the format to use when reading or writing Time fields to JSON or CSV.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format

Description

Specifies the format to use when reading or writing Timestamp fields to JSON or CSV.

Default Value

yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

true

Truststore Filename for Schema Registry

Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using HTTPS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Truststore Password for Schema Registry

Description

> The password used to access the contents of the truststore configured in the Truststore Filename property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type for Schema Registry

Description

> The type of the truststore configured in the Truststore Filename property.

Default Value

> JKS

Accepted Values

> BCFKS, PKCS12, JKS

Required

> true

HTTP Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the HTTP Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Basic Authentication Password

Description

> The password to be used for authentication when connecting to the URL specified in the Remote U RL property. If an authentication method other than Basic Authentication is used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

> false

Basic Authentication Username

Description

> The username used for authentication when connecting to the URL specified in the Remote U
> RL property. Cannot include control characters (0-31), ':', or DEL (127). If an authentication
> method other than Basic Authentication is used, this property must be completely removed from the
> configuration JSON.

Default Value

Accepted Values

Required

> false

Content-Type

Description

> The Content-Type to specify in the HTTP header of the POST request sent by this connector.

Default Value

> application/octet-stream

Accepted Values

Required

> true

Date Format

Description

> Specifies the format to use when reading date fields from JSON. If Forward Raw   Data is set to
> false, the format defined here also applies to the date fields in the output JSON message.

Default Value

> yyyy-MM-dd

Accepted Values

Required

> true

Forward Raw Data

Description

> Specifies whether messages from Kafka should be forwarded as is or converted to JSON. If set to
> false the Kafka Message Data Format parameter must be specified.

Default Value

> true

Accepted Values

> true, false

Required

> true

Kafka Message Data Format

Description

> Specifies the format of the messages the connector receives from Kafka. If the Forward Raw Data
> property is set to true then this property is ignored. However, even in a case like this, this property
> must be assigned a valid value.

Default Value

>AVRO

Accepted Values

>AVRO, JSON

Required

>true

Kerberos Keytab for Schema Registry

Description

>The fully-qualified filename of the Kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

>The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

>true

Kerberos Principal for Schema Registry

Description

>The Kerberos principal used for authenticating to Schema Registry.

Default Value

>default

Accepted Values

Required

>true

Keystore Filename for Secure HTTP

Description

>The fully-qualified filename of a keystore. This keystore is used to establish a secure connection with the HTTP server using mutual TLS.
>
>If the HTTP server does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

>false

Keystore Key Password for Secure HTTP

Description

>The password used to access the key stored in the keystore file configured in the Keystore Filename for Secure HTTP property.
>
>If the HTTP server does not require client certificate authentication, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

>   false

Keystore Password for Secure HTTP

Description

>   The password used to access the contents of the keystore configured in the Keystore Filename for
>   Secure HTTP property.
>
>   If the HTTP server does not require client certificate authentication, this property must be
>   completely removed from the configuration JSON.

Default Value

Accepted Values

Required

>   false

Keystore Type for Secure HTTP

Description

>   The type of the keystore configured in the Keystore Filename for Secure   HTTP property.
>
>   If the HTTP server does not require client certificate authentication, this property must be
>   completely removed from the configuration JSON.

Default Value

Accepted Values

>   BCFKS, PKCS12, JKS

Required

>   false

Output Grouping for JSON

Description

>   Specifies how JSON objects are grouped in the connector output.
>
>   If set to output-array, the output will consist of an array of JSON objects.
>
>   f set to output-oneline, each line of the output data becomes one JSON object. That is, each JSON
>   object occupies one line in the output.
>
>   This property is only used if forwarding raw data is disabled (Forward Raw Data is set to false).
>   This is because when forwarding raw data is disabled, the data coming from Kafka gets converted to
>   JSON. However, this property must still be set even if forwarding raw data is enabled.

Default Value

>   output-oneline

Accepted Values

>   output-array, output-oneline

Required

>   true

Remote URL

Description

The remote URL to connect to. Must include scheme, host, port, and path.

Default Value

https://localhost:22000/contentListener

Accepted Values

Required

true

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.

- If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
- If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
- If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
- If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.

This property is not used if record processing is disabled (Forward Raw   Data is set to true).

Default Value

Schema Registry

Accepted Values

Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry.

If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name.

If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Time Format

Description

Specifies the format to use when reading time fields from JSON. If Forward Raw   Data is set to false, the format defined here also applies to the time fields in the output JSON message.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format

Description

Specifies the format to use when reading timestamp fields from JSON. If Forward Raw Data is set to false, the format defined here also applies to the timestamp fields in the output JSON message.

Default Value

HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename for Schema Registry

Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using TLS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connection.

Accepted Values

Required

> true

Truststore Filename for Secure HTTP

Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with the HTTP server using TLS

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Truststore Password for Schema Registry

Description

> The password used to access the contents of the truststore configured in the Truststore Filename for Schema Registry property.

Default Value

> password

Accepted Values

Required

> true

Truststore Password for Secure HTTP

Description

> The password used to access the contents of the truststore configured in the Truststore Filename for Secure HTTP property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type for Schema Registry

Description

The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

Truststore Type for Secure HTTP

Description

The type of the truststore configured in the Truststore Filename for Secure   HTTP property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

InfluxDB Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the InfluxDB Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Batch Size

Description

Maximum size of records merged together and put to InfluxDB in one batch. Example values: 100 MB, 1 GB

Default Value

1 MB

Accepted Values

Required

true

Consistency Level

Description

Specifies the InfluxDB consistency level

Default Value

> ONE

Accepted Values

> ONRE ANY, ALL, QUORUM

Required

> true

Date Format

Description

> Specifies the format to use when reading date fields from JSON. If Forward Raw   Data is set to false, the format defined here also applies to the date fields in the output JSON message.

Default Value

> yyyy-MM-dd

Accepted Values

Required

> true

InfluxDB Connection URL

Description

> Specifies the InfluxDB URL to connect to.

Default Value

> http://localhost:8086

Accepted Values

Required

> true

InfluxDB Database Name

Description

> The name of the InfluxDB database to connect to.

Default Value

Accepted Values

Required

> true

InfluxDB User Name

Description

> The username for accessing InfluxDB.

Default Value

Accepted Values

Required

> false

InfluxDB User Password

Description

The password for InfluxDB user.

Default Value

Accepted Values

Required

false

Kafka Message Data Format

Description

Specifies the format of the messages the connector receives from Kafka. If the Forward Raw Data property is set to true then this property is ignored. However, even in a case like this, this property must be assigned a valid value.

Default Value

AVRO

Accepted Values

AVRO, JSON

Required

true

Kerberos Keytab for Schema Registry

Description

The fully-qualified filename of the Kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal for Schema Registry

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

default

Accepted Values

Required

true

Line Protocol Query

Description

A record schema based query which returns the attributes in line protocol format.

Default Value

Accepted Values

Required

true

Retention Policy

Description

Specifies the retention policy for saving records in InfluxDB.

Default Value

autogen

Accepted Values

Required

true

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.

- If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
- If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
- If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
- If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.

Default Value

Schema Registry

Accepted Values

Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry.

If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name.

If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Time Format

Description

Specifies the format to use when reading time fields from JSON. If Forward Raw   Data is set to false, the format defined here also applies to the time fields in the output JSON message.

Default Value

HH:mm:ss

Accepted Values

Required

true

Timestamp Format

Description

Specifies the format to use when reading timestamp fields from JSON. If Forward Raw Data is set to false, the format defined here also applies to the timestamp fields in the output JSON message.

Default Value

HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename for Schema Registry
Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using TLS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connection.

Accepted Values

Required

> true

Truststore Password for Schema Registry
Description

> The password used to access the contents of the truststore configured in the Truststore Filename for Schema Registry property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type for Schema Registry
Description

> The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

> JKS

Accepted Values

> BCFKS, PKCS12, JKS

Required

> true

JDBC Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the JDBC Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration

template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Database Connection URL

Description

The database specific connection URL used for connecting to the database. For example, jdbc:po stgresql://localhost:5432/postgres.

Default Value

Accepted Values

Required

true

Database Driver Class Name

Description

The database driver class name. For example, org.postgresql.Driver.

Default Value

Accepted Values

Required

true

Database Driver Location

Description

A comma-separated list of files or folders containing the JDBC client libraries.

Default Value

Accepted Values

Required

true

Database Schema Name

Description

The name of the database schema containing the table to load data.

Default Value

Accepted Values

Required

false

Database Table Name

Description

The name of the database table to load data.

Default Value

Accepted Values

Required

true

Database Type

Description

The database type used for generating database specific SQL queries.

Default Value

Generic

Accepted Values

Generic, Oracle, Oracle 12+, MS SQL 2008, MS SQL 2012+, MySQL, PostgreSQL

Required

true

Database User Name

Description

The database user name. If username/password authentication is not required by the database server, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Database User Password

Description

The database user password. If username/password authentication is not required by the database server, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Date Format

Description

Specifies the format to use when reading date fields from JSON.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Kafka Message Data Format

Description

Specifies the format of the messages the connector receives from Kafka.

Default Value

Avro

Accepted Values

Avro, JSON

Required

> true

Kerberos Keytab

Description

> The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

> The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Kerberos Principal

Description

> The Kerberos principal used for authenticating to Schema Registry.

Default Value

> default

Accepted Values

Required

> true

Quote Column Names

Description

> Specifies whether to use quotes around the column names in the generated SQL statement.

Default Value

> false

Accepted Values

> true, false

Required

> true

Quote Table Name

Description

> Specifies whether to use quotes around the table name in the generated SQL statement.

Default Value

> false

Accepted Values

> true, false

Required

> true

Schema Access Strategy

Description

Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.

- If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
- If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
- If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
- If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.

Default Value

Schema Registry

Accepted Values

Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

true

Schema Branch
Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name
Description

The schema name to look up in Schema Registry. If the Schema Access Strategy property is set to Schema Registry, this property must contain a valid schema name. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL
Description

The URL of the Schema Registry server. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

> true

Schema Version

Description

> The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Statement Type

Description

> Specifies the type of SQL statement that is generated to put data into the database.

Default Value

> INSERT

Accepted Values

> INSERT, INSERT_IGNORE, UPDATE, UPSERT

Required

> true

Time Format

Description

> Specifies the format to use when reading Time fields from JSON.

Default Value

> HH:mm:ss

Accepted Values

Required

> true

Timestamp Format

Description

> Specifies the format to use when reading Timestamp fields from JSON.

Default Value

> yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename

Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection
with Schema Registry using HTTPS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Truststore Password

Description

The password used to access the contents of the truststore configured in the Truststore Filename
property

Default Value

password

Accepted Values

Required

true

Truststore Type

Description

The type of the truststore configured in the Truststore Filename property.

Default Value

JKS

Accepted Values

BCFKS, PKCS12, JKS

Required

true

Kudu Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the Kudu Sink
connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect
framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the
SMM UI, all valid properties are presented in the default configuration template. You can view the configuration
template to get a full list of valid properties. In addition, for more information regarding the accepted properties not
listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Date Format

Description

Specifies the format to use when reading Date fields from JSON.

Default Value

> yyyy-MM-dd

Accepted Values

Required

> true

Handle Schema Drift

Description

> Specifies whether to handle Schema Drift. If set to true, when fields with names that are not in the
> target Kudu table are encountered, the Kudu table is altered to include new columns for those fields.
> If set to false, fields that only exist in the input data are skipped.

Default Value

> false

Accepted Values

> true, false

Required

> true

Ignore Null

Description

> Ignore NULL on Kudu Put Operation. If set to true, only non-null columns get updated.

Default Value

> false

Accepted Values

> true, false

Required

> true

Kafka Message Data Format

Description

> Specifies the format of the messages the connector receives from Kafka.

Default Value

> AVRO

Accepted Values

> AVRO, JSON

Required

> true

Kerberos Keytab for Kudu

Description

> The fully-qualified filename of the kerberos keytab associated with the principal for accessing
> Kudu.

Default Value

> The location of the default keytab which is empty and can only be used for unsecure connections

Accepted Values

Required

> true

Kerberos Keytab for Schema Registry

Description

> The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

> The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Kerberos Principal for Kudu

Description

> The Kerberos principal used for authenticating to Kudu.

Default Value

> default

Accepted Values

Required

> true

Kerberos Principal for Schema Registry

Description

> The Kerberos principal used for authenticating to Schema Registry.

Default Value

> default

Accepted Values

Required

> true

Kudu Kerberos Service Name

Description

> The SASL protocol name to use for authenticating with Kerberos. Must match the Kudu service principal name.

Default Value

> kudu

Accepted Values

Required

> false

Kudu Masters

Description

> The comma separated addresses of the Kudu masters to connect to. For example: localhost: 7051 ,localhost:7151,localhost:7251.

Default Value

Accepted Values

Required

> true

Kudu Operation Type

Description

> Specifies what Kudu operation gets executed.

Default Value

> INSERT

Accepted Values

> INSERT, INSERT_IGNORE, UPSERT, UPDATE, DELETE, UPDATE_IGNORE,
> DELETE_IGNORE

Required

> true

Lowercase Field Names

Description

> Specifies whether to convert column names to lowercase when finding indexes of Kudu table
> columns.

Default Value

> false

Accepted Values

> true, false

Required

> true

Schema Access Strategy

Description

> Specifies the strategy used for determining the schema of the Kafka record. The value you set here
> depends on the data format set in Kafka Message Data   Format.
>
> - If set to Schema Registry, the schema is read from Schema Registry. This setting can be used
>   with both Avro and JSON formats.
> - If set to Infer Schema, the schema is inferred based on the input file. This setting can only be
>   used if Kafka Message Data Format is JSON.
> - If set to Embedded Schema, the schema embedded in the input is used. This setting can only be
>   used if Kafka Message Data Format is Avro.
> - If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry.
>   This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro
>   messages are expected to have a reference to the schema in Schema Registry encoded within the
>   message content.

Default Value

> Schema Registry

Accepted Values

> Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

true

Schema Branch

Description

The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Schema Name

Description

The schema name to look up in Schema Registry. If the Schema Access   Strategy property is set to Schema Registry, this property must contain a valid schema name. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

Accepted Values

Required

false

Schema Registry URL

Description

The URL of the Schema Registry server. If Schema Registry is not used, use the default value.

Default Value

http://localhost:7788/api/v1

Accepted Values

Required

true

Schema Version

Description

The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

false

Table Name With Schema

Description

The schema and name of the Kudu Table to put data into. For example: default.mytable. When using Impala: impala::default.mytable.

Default Value

Accepted Values

Required

true

Time Format
Description

Specifies the format to use when reading Time fields from JSON.

Default Value

HH:mm:ss

Accepted Values

Required

true

Truststore Filename for Schema Registry
Description

The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using TLS.

Default Value

The location of the default truststore which is empty and can only be used for unsecure connections

Accepted Values

Required

true

Truststore Password for Schema Registry
Description

The password used to access the contents of the truststore configured in the Truststore Filename for Schema Registry property.

Default Value

password

Accepted Values

Required

true

Truststore Type for Schema Registry
Description

The type of the truststore configured in the Truststore Filename for Schema   Registry property.

Default Value

Accepted Values

BCFKS, PKCS12, JKS

Required

true

S3 Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the S3 Sink connector.

The properties listed in this reference must be added to the connector configuration with the following prefix:

```
parameter.[***CONNECTOR NAME***] Parameters:
```

In addition to the properties listed here, this connector also accepts certain properties of the Kafka Connect framework as well as the properties of the NiFi Stateless Sink connector. When creating a new connector using the SMM UI, all valid properties are presented in the default configuration template. You can view the configuration template to get a full list of valid properties. In addition, for more information regarding the accepted properties not listed here, you can review the Apache Kafka documentation and the Stateless NiFi Sink properties reference.

Avro Schema Write Strategy

Description

Specifies whether the record schema is written to the output data file. Applicable only for Avro output (Output File Data Format is set to Avro).

Default Value

Embed Avro Schema

Accepted Values

Embed Avro Schema, Do Not Write Schema, HWX Content-Encoded Schema Reference

Required

false

AWS Access Key ID

Description

The Access Key ID to use for authentication to AWS.

Default Value

Accepted Values

Required

true

AWS Secret Access Key

Description

The Secret Access Key to use for authentication to AWS.

Default Value

Accepted Values

Required

true

Date Format

Description

Specifies the format to use when writing date fields to JSON or CSV.

Default Value

yyyy-MM-dd

Accepted Values

Required

true

Kafka Message Data Format

Description

Specifies the format of the messages the connector receives from Kafka. If set to Avro or JSON, record processing is enabled. Raw can be used for unstructured text or binary data

Default Value

Avro

Accepted Values

Avro, JSON, Raw

Required

true

Kerberos Keytab

Description

The fully-qualified filename of the kerberos keytab associated with the principal for accessing Schema Registry.

Default Value

The location of the default keytab which is empty and can only be used for unsecure connections.

Accepted Values

Required

true

Kerberos Principal

Description

The Kerberos principal used for authenticating to Schema Registry.

Default Value

default

Accepted Values

Required

true

Maximum File Size

Description

The maximum size of the output data file. No size limit is applied if this property is not specified. Example values: 100 MB, 1 GB.

Default Value

Accepted Values

Required

false

Maximum Number of Entries

Description

The maximum number of entries in the output data file. In the context of this property, entry can mean one of two things. If record processing is enabled (Kafka Message Data Format is set to Avro or JSON), an entry is a record. Otherwise, entry means a Kafka message. Set this property to 1 if you want to create a separate output file for each Kafka message.

Default Value

1000000

Accepted Values

Required

true

Output File Data Format

Description

Specifies the format of the records written to the output file. Required when record processing is enabled (Kafka Message Data Format is set to Avro or JSON).

Default Value

Avro

Accepted Values

Avro, JSON, CSV, Parquet

Required

Output File Demarcator

Description

Specifies the character sequence for demarcating (delimiting) message boundaries when multiple Kafka messages are ingested into an output file as raw messages (no record processing). This property can only be used if Kafka Message Data   Format is set to Raw. If you want to use newline as the delimiter, set this property to \n.

Default Value

Accepted Values

Required

false

Output Filename Pattern

Description

Specifies the structure of the name of the output file (S3 object name including the prefix). The pattern can contain string literal (fixed text) parts and one or more of the following expressions:

- ${filename.uuid}: Inserts a generated UUID in the filename.
- ${filename.timestamp}: Inserts the current timestamp in the filename.
- ${filename.sequence}: Inserts an incrementing sequence value in the filename.

In order to generate unique filenames, either ${filename.uuid} or ${filename.sequence} must be used in the pattern.

Examples:

- data_${filename.uuid}.json
- records_${filename.timestamp}_${filename.sequence}.avro

Default Value

> ${filename.uuid}

Accepted Values

Required

> false

Output Filename Sequence Initial Value

Description

> This property is used to configure the initial value of the ${filename.sequence} expression. The value you set in this property is not the initial value of the sequence. The sequence starts at the value of this property +1. For example, if you set this property to 0, the sequence starts at 1.

Default Value

> 0

Accepted Values

Required

> false

Output Filename Sequence Padding Length

Description

> Specifies the length of the ${filename.sequence} expression in characters. If the sequence has fewer characters than the value set in this property, it is padded with zeros (0). Padding is added to the left of the sequence.

Default Value

> 6

Accepted Values

Required

> false

Output Filename Timestamp Format

Description

> The timestamp format to use for the ${filename.timestamp} expression. For example, yyyyMMdd _HHmmss_SSS.

Default Value

Accepted Values

Required

> false

Parquet Compression Type

Description

> The type of compression used for writing parquet files. Required when Output   File Data Format is Parquet.

Default Value

> UNCOMPRESSED

Accepted Values

> UNCOMPRESSED, SNAPPY, GZIP, LZO, BROTLI, LZ4, ZSTD

Required

> false

S3 Bucket

Description

> The name of the S3 Bucket where the output file is uploaded.

Default Value

Accepted Values

Required

> false

S3 Region

Description

> The Region of the S3 Bucket.

Default Value

Accepted Values

Required

> true

S3 Storage Class

Description

> The Storage Class to use for the output file on S3.

Default Value

> Standard

Accepted Values

> Standard, IntelligentTiering, StandardInfrequentAccess, OneZoneInfrequentAccess, Glacier, GlacierInstantRetrieval, DeepArchive, ReducedRedundancy

Required

> true

Schema Access Strategy

Description

> Specifies the strategy used for determining the schema of the Kafka record. The value you set here depends on the data format set in Kafka Message Data   Format.
>
> - If set to Schema Registry, the schema is read from Schema Registry. This setting can be used with both Avro and JSON formats.
> - If set to Infer Schema, the schema is inferred based on the input file. This setting can only be used if Kafka Message Data Format is JSON.
> - If set to Embedded Schema, the schema embedded in the input is used. This setting can only be used if Kafka Message Data Format is Avro.
> - If set to HWX Content-Encoded Schema Reference, the schema is read from Schema Registry. This setting can only be used if Kafka Message   Data Format is Avro. In this case the Avro messages are expected to have a reference to the schema in Schema Registry encoded within the message content.
>
> This property is not used if record processing is disabled (Kafka   Message Data Format is set to Raw).

Default Value

> Schema Registry

Accepted Values

> Schema Registry, Infer Schema, Embedded Schema, HWX Content-Encoded Schema Reference

Required

> true

Schema Branch

Description

> The name of the branch to use when looking up the schema in Schema Registry. If Schema Registry is not used, this property must be completely removed from the configuration JSON

Default Value

Accepted Values

Required

> false

Schema Name

Description

> The name of the branch to use when looking up the schema in Schema Registry. Schema Branch and Schema Version cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Schema Registry URL

Description

> The URL of the Schema Registry server. If Schema Registry is not used, this property must be completely removed from the configuration JSON.

Default Value

> http://localhost:7788/api/v1

Accepted Values

Required

> true

Schema Version

Description

> The version of the schema to look up in Schema Registry. If Schema Registry is used and a schema version is not specified, the latest version of the schema is retrieved. Schema Branch and Schema V ersion cannot be specified at the same time. If one is specified, the other needs to be removed from the configuration. If Schema Registry is not used, this property must be completely removed from the configuration.

Default Value

Accepted Values

Required

> false

Time Format

Description

> Specifies the format to use when writing Time fields to JSON or CSV.

Default Value

> HH:mm:ss

Accepted Values

Required

> true

Timestamp Format

Description

> Specifies the format to use when writing Timestamp fields to JSON or CSV.

Default Value

> yyyy-MM-dd HH:mm:ss.SSS

Accepted Values

Required

> true

Truststore Filename

Description

> The fully-qualified filename of a truststore. This truststore is used to establish a secure connection with Schema Registry using HTTPS.

Default Value

> The location of the default truststore which is empty and can only be used for unsecure connections.

Accepted Values

Required

> true

Truststore Password

Description

> The password used to access the contents of the truststore configured in the Truststore Filename property.

Default Value

> password

Accepted Values

Required

> true

Truststore Type

Description

> The type of the truststore configured in the Truststore Filename property.

Default Value

> JKS

Accepted Values

> BCFKS, PKCS12, JKS

Required

> true

Stateless NiFi Sink properties reference

Review the following reference for a comprehensive list of the connector properties that are specific to the Stateless NiFi Sink connector.

In addition to the properties listed here, this connector also accepts the properties of the Kafka Connect framework. For a comprehensive list of these properties, see the Apache Kafka documentation.

attribute.prefix
Description

> The prefix to add to the key of each header that matches the regular expression specified in head ers.as.attributes.regex. For example, if the header key is MyHeader, its value is MyValue, headers. as.attributes.regex is set to My.*, and this property is set to kafka, the flowfile that is created for the Kafka message will have an attribute named kafka.MyHeader with a value of MyValue.

Default Value

Accepted Values

Required

> false

dataflow.timeout
Description

> Specifies the maximum amount of time to wait for the dataflow to complete. If the dataflow does not complete before this timeout, the thread is interrupted and the dataflow is considered as a failure. The session is rolled back and the connector retriggers the flow. Defaults to 60 seconds if not specified.

Default Value

> 60 seconds

Accepted Values

Required

> false

extensions.directory
Description

> Specifies the directory that stores downloaded extensions. Extensions are the NAR (NiFi Archive) files containing the processors and controller services a flow might use. Since Stateless NiFi is only the NiFi engine, it does not contain any of the processors and controller services you might use in your flow. When deploying the connector with the custom flow, the system needs to download the specific extensions that your flow uses from Nexus (unless they are already present in this directory). These extensions are stored in this directory. Because the default directory might not

be writable, and to aid in upgrade scenarios, Cloudera recommends that you always specify an
extensions directory.

Default Value

/tmp/nifi-stateless-extensions

Accepted Values

Required

true

failure.ports

Description

A comma separated list of output ports that are considered as failure conditions. If any flowfile
is routed to an output port specified in this property, the dataflow is considered a failure and the
session is rolled back. After a set amount of time, the dataflow reattempts to process the Kafka
record. Any data transferred to an output port that is not in the list of failure ports is discarded.

Because of how Stateless NiFi Sink connectors behave, even if a single flowfile ends up in an
output port that is marked as failure, the entire sessions is rolled back with all messages in the batch.
Furthermore, if a flowfile ends up in a failure port in each subsequent iteration, the result is an
endless loop. With some sink connectors (for example. MQTT Sink) this is the desired behavior.
For more information regarding this behavior, see Dataflow execution and scheduling.

Default Value

Accepted Values

Required

false

flow.snapshot

Description

Specifies the dataflow to run. When using SMM to deploy a connector, the value you set in this
property must be a JSON object. URLs, file paths, or escaped JSON strings are not supported when
using SMM. Alternatively, if using the Kafka Connect REST API to deploy a connector, this can
be a file containing the dataflow, a URL that points to a dataflow, or a string containing the entire
dataflow as an escaped JSON. Cloudera however, does not recommend using the Kafka Connect
REST API to interact with this connector or Kafka Connect.

Default Value

Accepted Values

Required

true

headers.as.attributes.regex

Description

A Java regular expression that is evaluated against all Kafka record headers. Headers are added to
the flowfile as an attribute if the header key matches the regular expression. The header key is used
as the attribute name. The header value is used as the attribute value. Additionally, the name of the
attribute can also contain an optional prefix which is defined by the attribute.prefix property.

Default Value

Accepted Values

Required

false

input.port

Description

> The name of the input port in the NiFi dataflow that Kafka records are sent to. If the dataflow contains exactly one input port, this property is optional and can be omitted. However, if the dataflow contains multiple input ports, this property must be specified.

Default Value

Accepted Values

Required

> false

krb5.file

Description

> Specifies the krb5.conf file to use if the dataflow interacts with any services that are secured using Kerberos. Defaults to /etc/krb5.conf if not specified.

Default Value

> /etc/krb5.conf

Accepted Values

Required

> false

name

Description

> The name of the connector. On the SMM UI, the connector names are specified using the Enter Name field. The name that you enter in the Enter Name field is automatically set as the value of the name property when the connector is deployed. Because of this, the name property is omitted from the configuration template provided in SMM. If you manually add the name property to the configuration in SMM, ensure that the value you set matches the connector name specified in the Enter Name field. Otherwise, the connector fails to deploy.

Default Value

Accepted Values

Required

> true

nexus.url

Description

> Specifies the Base URL of the Nexus instance to source extensions from. If configuring a Nexus instance that has multiple repositories, include the name of the repository in the URL. For example, https://nexus-private.myorganization.org/nexus/repository/my-repository/. If the property is not specified, the necessary extensions (the ones used by the flow) must be provided in the extensions directory before deploying the connector.

Default Value

Accepted Values

Required

> true

parameter.[***FLOW PARAMETER NAME***]

Description

> Specifies a parameter to use in the dataflow. For example, assume that you have the following entry in your connector configuration "parameter.Directory":   "/mydir". In a case like this, any parameter context in the dataflow that has a parameter named Directory gets the specified value (/mydir). If the dataflow has child process groups, and those child process groups have their own parameter contexts, the value is used for all parameter contexts that contain a parameter named Directory. Parameters can also be applied to specific parameter contexts only. This can be done by prefixing the parameter name (Directory) with the name of the parameter context followed by a colon. For example, parameter.My Context:Directory only applies the specified value for the Dire ctory parameter in the Parameter Context named My Context.

Default Value

Accepted Values

Required

> false

working.directory

Description

> Specifies a directory on the Connect server that NiFi should use for unpacking extensions that it needs to perform the dataflow. The contents of extensions.directory are unpacked here. Defaults to /tmp/nifi-stateless-working if not specified.

Default Value

> /tmp/nifi-stateless-working

Accepted Values

Required

> false