# MQTT filter to Kafka

**Date published: 2021-04-06**
**Date modified: 2025-09-30**

## CLOUDERA

# Legal Notice

# Contents

# ReadyFlow: MQTT filter to Kafka

You can use the MQTT filter to Kafka ReadyFlow to move your data from an MQTT source to a Kafka topic.

This ReadyFlow consumes JSON, CSV or Avro data from a source MQTT topic and parses the schema by looking up the schema name in the Cloudera Schema Registry. You can filter events by specifying a SQL query in the 'Filter Rule' parameter. The filtered events are then converted to the specified output data format and written to the destination Kafka topic. Failed Kafka write operations are retried automatically to handle transient issues. Define a KPI on the failure_WriteToKafka connection to monitor failed write operations.

**Note:** The default filter criteria allows all records to pass through.

| ReadyFlow details | |
|---|---|
| Source | MQTT Processor |
| Source Format | JSON, CSV, Avro |
| Destination | Kafka Topic |
| Destination Format | JSON, CSV, Avro |

### Moving data with MQTT filter to Kafka flow

MQTT is a messaging protocol for the Internet of Things (IoT) that is used in a wide variety of industries. You can use the MQTT filter to Kafka ReadyFlow to read data from a MQTT topic, parse the data using a schema provided either through the Cloudera Schema Registry or through direct schema input. You can then filter or transform the data using a SQL query before it is written to a destination topic in a Cloudera Kafka cluster.

# Prerequisites

Learn how to collect the information you need to deploy the MQTT filter to Kafka ReadyFlow, and meet other prerequisites.

### For your data ingest source

- You have the MQTT Broker URI.
- You have the MQTT Broker Username and Password to authenticate to the MQTT Broker.
- You have the MQTT Topic Filter to designate which source topics to subscribe to.

### For Cloudera Data Flow

- You have enabled Cloudera Data Flow for an environment.

  For information on how to enable Cloudera Data Flow for an environment, see Enabling Cloudera Data Flow for an Environment.
- You have created a Machine User to use as the Cloudera Workload User.

- You have given the Cloudera Workload User the EnvironmentUser role.

    1. From the Management Console, go to the environment for which Cloudera Data Flow is enabled.
    2. From the Actions drop down, click Manage Access.
    3. Identify the user you want to use as a Workload User.

        **Note:**

        The Cloudera Workload User can be a machine user or your own user name. It is best practice to create a dedicated Machine user for this.
    4. Give that user EnvironmentUser role.
- You have synchronized your user to the Cloudera on cloud environment that you enabled for Cloudera Data Flow.

    For information on how to synchronize your user to FreeIPA, see Performing User Sync.
- You have granted your Cloudera user the DFCatalogAdmin and DFFlowAdmin roles to enable your user to add the ReadyFlow to the Catalog and deploy the flow definition.

    1. Give a user permission to add the ReadyFlow to the Catalog.

        a. From the Management Console, click User Management.
        b. Enter the name of the user or group you wish to authorize in the Search field.
        c. Select the user or group from the list that displays.
        d. Click  Roles Update Roles .
        e. From Update Roles, select DFCatalogAdmin and click Update.

            **Note:** If the ReadyFlow is already in the Catalog, then you can give your user just the DFCatalogViewer role.

    2. Give your user or group permission to deploy flow definitions.

        a. From the Management Console, click Environments to display the Environment List page.
        b. Select the environment to which you want your user or group to deploy flow definitions.
        c. Click  Actions Manage Access  to display the Environment Access page.
        d. Enter the name of your user or group you wish to authorize in the Search field.
        e. Select your user or group and click Update Roles.
        f. Select DFFlowAdmin from the list of roles.
        g. Click Update Roles.

    3. Give your user or group access to the Project where the ReadyFlow will be deployed.

        a. Go to  Data Flow Projects .
        b. Select the project where you want to manage access rights and click  ⋮  More Manage Access .

    4. Start typing the name of the user or group you want to add and select them from the list.
    5. Select the Resource Roles you want to grant.
    6. Click Update Roles.
    7. Click Synchronize Users.

## For your data ingest target

- You have created a Streams Messaging cluster in Cloudera on cloud to host your Schema Registry.

    For information on how to create a Streams Messaging cluster, see Setting up your Streams Messaging Cluster.

- You have created at least one Kafka topic.

  1. Navigate to Management Console > Environments and select your environment.
  2. Select your Streams Messaging cluster.
  3. Click on the Streams Messaging Manager icon.
  4. Navigate to the Topics page.
  5. Click Add New and provide the following information:
     - Topic name
     - Number of partitions
     - Level of availability
     - Cleanup policy

       **Tip:**
       SMM has automatically set Kafka topic configuration parameters. To manually adjust them, click Advanced.
  6. Click Save.
- You have created a schema for your data and have uploaded it to the Schema Registry in the Streams Messaging cluster.

  For information on how to create a new schema, see Creating a new schema. For example:

```
{
    "type":"record",
    "name":"SensorReading",
    "namespace":"com.cloudera.example",
    "doc":"This is a sample sensor reading",
    "fields":[
        {
            "name":"sensor_id",
            "doc":"Sensor identification number.",
            "type":"int"
        },
        {
            "name":"sensor_ts",
            "doc":"Timestamp of the collected readings.",
            "type":"long"
        },
        {
            "name":"sensor_0",
            "doc":"Reading #0.",
            "type":"int"
        },
        {
            "name":"sensor_1",
            "doc":"Reading #1.",
            "type":"int"
        },
        {
            "name":"sensor_2",
            "doc":"Reading #2.",
            "type":"int"
        },
        {
            "name":"sensor_3",
            "doc":"Reading #3.",
            "type":"int"
        }
    ]
```

```
}
```

- You have the Schema Registry Host Name.

  1. From the Management Console, go to Data Hub Clusters and select the Streams Messaging cluster you are using.
  2. Navigate to the **Hardware** tab to locate the Master Node FQDN. Schema Registry is always running on the Master node, so copy the Master node FQDN.

- You have the Kafka broker end points.

  1. From the Management Console, click Data Hub Clusters.
  2. Select the Streams Messaging cluster from which you want to ingest data.
  3. Click the Hardware tab.
  4. Note the Kafka Broker FQDNs for each node in your cluster.
  5. Construct your Kafka Broker Endpoints by using the FQDN and Port number 9093 separated by a colon. Separate endpoints by a comma. For example:

     ```
     broker1.fqdn:9093,broker2.fqdn:9093,broker3.fqdn:9093
     ```

     Kafka broker FQDNs are listed under the **Core_broker** section.

- You have the Kafka Consumer Group ID.

  This ID is defined by the user. Pick an ID and then create a Ranger policy for it. Use the ID when deploying the flow in Cloudera Data Flow.

- You have assigned the Cloudera Workload User policies to access the consumer group ID and topic.

  1. Navigate to Management Console > Environments, and select the environment where you have created your cluster.
  2. Select Ranger. You are redirected to the Ranger **Service Manager** page.
  3. Select your Streams Messaging cluster under the **Kafka** folder.
  4. Create a policy to enable your Workload User to access the Kafka source topic.
  5. On the **Create Policy** page, give the policy a name, select topic from the drop-down list, add the user, and assign the Consume permission.
  6. Create another policy to give your Workload User access to the consumer group ID.
  7. On the **Create Policy** page, give the policy a name, select consumergroup from the drop-down list, add the user, and assign the Consume permission.

- You have assigned the Cloudera Workload User read-access to the schema.

  1. Navigate to  Management Console Environments , and select the environment where you have created your cluster.
  2. Select Ranger. You are redirected to the Ranger **Service Manager** page.
  3. Select your Streams Messaging cluster under the **Schema Registry** folder.
  4. Click Add New Policy.
  5. On the **Create Policy** page, give the policy a name, specify the schema details, add the user, and assign the Read permission.

**Related Concepts**
Required parameters

# Required parameters

When deploying the MQTT filter to Kafka ReadyFlow, you have to provide the following parameters. Use the information you collected in *Prerequisites*.

**Table 1: MQTT filter to Kafka ReadyFlow configuration parameters**

| Parameter Name | Description |
|---|---|
| CDP Workload User | Specify the Cloudera machine user or workload user name that you want to use to authenticate to Kafka. Ensure this user has the appropriate access rights to the destination Kafka topic. |
| CDP Workload User Password | Specify the password of the Cloudera machine user or workload user you are using to authenticate against Kafka. |
| CSV Delimiter | If your source data is CSV, specify the delimiter here. |
| Data Input Format | Specify the format of your input data. Supported values are:<br>• CSV<br>• JSON<br>• AVRO |
| Data Output Format | Specify the desired format for your output data. Supported values are:<br>• CSV<br>• JSON<br>• AVRO |
| Filter Rule | Specify the filter rule expressed in SQL to filter streaming events for the destination topic. Records matching the filter will be written to the destination topic. The default value forwards all records. |
| Kafka Broker Endpoint | Specify the Kafka bootstrap servers string as a comma separated list. |
| Kafka Destination Topic | Specify the destination topic name. |
| Kafka Producer ID | Specify the id for the Kafka producer writing into the destination topic. |
| MQTT Broker Password | Specify the password to use when connecting to the MQTT broker. |
| MQTT Broker URI | Specify the URI to use to connect to the MQTT broker. |
| MQTT Broker Username | Specify the username to use when connecting to the MQTT broker. |
| MQTT Topic Filter | Specify the MQTT topic filter to designate the topics to subscribe to. |
| Schema Name | Specify the schema name to be looked up in the Schema Registry. |
| Schema Registry Hostname | Specify the hostname of the Schema Registry you want to connect to. This must be the direct hostname of the Schema Registry itself, not the Knox Endpoint. |

**Related Concepts**
Prerequisites
**Related Information**
Deploying a ReadyFlow