Cloudera DataFlow for Data Hub 7.2.14

Connecting Kafka Clients to Data Hub Provisioned Clusters

Date published: 2022-02-24
Date modified: 2022-02-24

https://docs.cloudera.com/

Legal Notice

Contents

Connecting Kafka clients to Data Hub provisioned clusters

Learn how to connect Kafka clients to clusters provisioned with Data Hub.

About this task

Use the following steps to connect Kafka clients to clusters provisioned with Data Hub. Configuration examples provided in this list of steps assume that the cluster you are connecting to was provisioned with a Streams Messaging cluster definition.

Note:

These instructions are for connecting Kafka clients to Data Hub provisioned Kafka clusters. For information on connecting NiFi to Kafka within the same CDP Public Cloud environment, see Ingesting data into Apache Kafka.

Before you begin

- If you are connecting your clients from outside of your virtual network (VPC or Vnet) verify that both inbound and outbound traffic is enabled on the port used by Kafka brokers for secure communication. The default port is 9093. For more information, see the following resources:
  - AWS: Security Groups for Your VPC
  - Azure: How to open ports to a virtual machine with the Azure portal
- If you are connecting your clients over the internet, verify that your virtual network (VPC or Vnet) is assigned a public IP address. For more information, see the following resources:
  - AWS: IP Addressing in Your VPC
  - Azure: Associate a public IP address to a virtual machine
- Clients connecting to Data Hub provisioned clusters require a CDP user account that provides access to the required CDP resources. Verify that a CDP user account with the required roles and permissions is available for use. If not, create one. Any type of CDP user account can be used. If you are creating a new account to be used by Kafka clients, Cloudera recommends that you create a machine user account. For more information, see User Management in the Cloudera Management Console documentation.
- In addition to the CDP user account having access to the required CDP resources, the user account also needs to have the correct policies assigned to it in Ranger. Otherwise, the client cannot perform tasks on Kafka resources. These policies are specified within the Ranger instance that provides authorization to the Kafka service you want to connect to. For more information, see the Cloudera Runtime documentation on Apache Ranger and the Kafka specific Ranger documentation.

Procedure

1. Obtain the FreeIPA certificate of your environment:
   a) From the CDP Home Page navigate to Management Console Environments .
   b) Locate and select your environment from the list of available environments.
   c) Go to the FreeIPA tab.
   d) Click Get FreeIPA Certificate.
   The FreeIPA certificate file, *[***ENVIRONMENT NAME***]*.crt, is downloaded to your computer.

2. Add the FreeIPA certificate to the truststore of the client.

   The certificate needs to be added for all clients that you want to connect to the Data Hub provisioned cluster. The exact steps of adding the certificate to the truststore depends on the platform and key management software used. For example, you can use the Java Keytool command line tool:

   ```
   keytool -import -keystore [***CLIENT TRUSTSTORE.JKS***] -al
   ias [***ALIAS***] -file [***FREEIPA CERTIFICATE***]
   ```

   

   Tip: This command creates a new truststore file if the file specified with the -keystore option does not exist.

3. Obtain CDP workload credentials:

   A valid workload username and password has to be provided to the client, otherwise it cannot connect to the cluster. Credentials can be obtained from Management Console.

   a) From the CDP Home Page navigate to Management Console User Management.
   b) Locate and select the user account you want to use from the list of available accounts.
      The user details page displays information about the user.
   c) Find the username found in the Workload Username entry and note it down.
   d) Find the Workload Password entry and click Set Workload Password.
   e) In the dialog box that appears, enter a new workload password, confirm the password and note it down.
   f) Fill out the Environment text box.
   g) Click Set Workload Password and wait for the process to finish.
   h) Click close.

4. Configure clients.

   In order for clients to be able to connect to Kafka brokers, all required security related properties have to be added to the client's properties file. The following example configuration lists the default properties that are needed when connecting clients to a cluster provisioned by Data Hub with a Streams Messaging cluster definition. If you made changes to the security configuration of the brokers, or provisioned a custom cluster with non-default Kafka security settings, make sure to change the appropriate parameters in the client configuration as well.

   ```
   security.protocol=SASL_SSL
   sasl.mechanism=PLAIN
   ssl.truststore.location=[***CLIENT TRUSTSTORE.JKS***]
   ssl.truststore.password=[***TRUSTSTORE PASSWORD***]
   sasl.jaas.config=org.apache.kafka.common.security.plain.PlainLoginModule
    required \
       username="[***USERNAME***]" \
       password="[***PASSWORD***]";
   ```

   Replace *[***CLIENT TRUSTSTORE.JKS***]* with the path to the client's truststore file.This is the same file that you added the FreeIPA certificate to in Step 2 on page 5.

   Replace *[***TRUSTSTORE PASSWORD***]* with the password of the truststore file.

   Replace *[***USERNAME***]* and *[***PASSWORD***]* with the workload username and password obtained in Step 3 on page 5.

5. Obtain Kafka broker hostnames:

   You can obtain the Kafka broker hostnames from the Cloudera Manager UI.

   a) From the CDP Home Page navigate to Management Console Environments .
   b) Locate and select your environment from the list of available environments.
   c) Select the Data Hub cluster you want to connect to from the list of available clusters.
   d) Click the link found in the Cloudera Manger Info section.
      You are redirected to the Cloudera Manager web UI.
   e) Click Clusters and select the cluster that the Kafka service is running on.

      The default name for clusters created with a Streams Messaging Cluster definition is streams-messaging.
   f) Select the Kafka service.
   g) Go to Instances.

      The Kafka broker hostnames are listed in the Hostname column.

6. Connect clients to brokers.

   Connect the clients by supplying them with the broker hostnames obtained in step 5 on page 6. The actions you need to take differ depending on the type of client you are using.

   Custom developed Kafka Applications

   When producing or consuming messages with your own Kafka client application, you have to provide the Kafka broker hostnames within the client code.

   Kafka console producer and consumer

   When producing or consuming messages with the kafka-console-consumer or kafka-console-producer command line tools, run the producer or consumer with the appropriate hostnames. Additionally, you must also pass the client properties file containing the security related properties with --producer.config or --consumer.config. For example:

   ```
   kafka-console-producer --broker-
   list [***HOSTNAME***]:[***PORT***] --topic [***TOPIC***] --produ
   cer.config [***CLIENT PROPERTIES FILE***]
   ```

   ```
   kafka-console-consumer --bootstrap-serve
   r [***HOSTNAME***]:[***PORT***] --topic [***TOPIC***] --from-begi
   nning --consumer.config [***CLIENT PROPERTIES FILE***]
   ```

   Replace *[***CLIENT PROPERTIES FILE***]* with the path to the client's properties file. This is the same file that you updated in Step 4 on page 5.

7. To connect to Schema Registry, you must also set the following properties:

   ```
   schema.registry.url=<SCHEMA_REGISTRY_ENDPOINT>
   schema.registry.auth.username=<USERNAME>
   schema.registry.auth.password=<PASSWORD>
   ```

   - If schema.registry.url is not set, the client does not try to connect to Schema Registry.
   - If this property is set, the username and password must also be configured.
   - The Schema Registry endpoint can be found in the CDP Console, under the Endpoints tab of your DataHub Kafka cluster.

   Some applications may need to connect directly to Schema Registry to interact with it. For example, to retrieve or store a schema.

   A Schema Registry API client is also provided by Cloudera for these cases. For example, to retrieve the latest version of a schema you can use the following:

   ```
   Map<String, Object> config = new HashMap<>();
   config.put("schema.registry.url", "<SCHEMA_REGISTRY_ENDPOINT>");
   config.put("schema.registry.auth.username", "<USERNAME>");
   ```

```
config.put("schema.registry.auth.password", "<PASSWORD>");

String topicName = "my-topic";
SchemaRegistryClient client = new SchemaRegistryClient(config);
try {
    SchemaVersionInfo latest = client.getLatestSchemaVersionInfo(topicNam
e);
    System.out.println(latest.getSchemaText());
} catch (SchemaNotFoundException e) {
    LOG.info("Schema [{}] not found", topicName);
    throw e;
}
```

Results

Kafka clients and Schema Registry are configured and are able to connect to Data Hub provisioned clusters.