

Cloudera Runtime 7.2.10

Use cases for Streams Replication Manager in CDP Public Cloud

Date published: 2021-03-04

Date modified: 2021-06-08

CLouDERA

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

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

Using Streams Replication Manager in CDP Public Cloud overview.....	4
Replicating data from CDP PvC Base cluster to Data Hub cluster with SRM running in CDP PvC Base cluster.....	5
Replicating data from CDP PvC Base cluster to Data Hub cluster with SRM deployed in Data Hub cluster.....	11
Replicating data between Data Hub clusters with SRM deployed in a Data Hub cluster.....	17

Using Streams Replication Manager in CDP Public Cloud overview

Streams Replication Manager (SRM) can be deployed in both CDP PvC Base (on-prem) and Data Hub (cloud) clusters. You can use your SRM deployment to replicate Kafka data between CDP PvC Base and Data Hub clusters, or to replicate data between multiple Data Hub clusters. Review the following information to learn more about your deployment options, as well as the prerequisites and use cases for using SRM in a cloud-based context.

Starting with the December 2020 release of CDP Public Cloud, SRM is included in the default Streams Messaging cluster definitions. As a result, you can deploy SRM in a Data Hub cluster and use it to replicate data between all types of CDP clusters. This includes deployments with either Data Hub, CDP PvC Base, or both.

The following sections provide information on how you can deploy SRM in a Data Hub cluster, what prerequisites you must meet before using SRM, and the common use cases where you can use SRM in a cloud-based context.

Differences between light and heavy deployments

In CDP Public Cloud, SRM can be deployed in Data Hub clusters with both the light and heavy duty variants of the Streams Messaging cluster definition. However, there are significant differences in how SRM is deployed with each definition:

Light duty definition:

In the light duty definition, SRM is deployed by default on the broker and master hosts of the cluster. This means that SRM is available for use by default in a Data Hub cluster provisioned with the light duty definition.

Heavy duty definition

In the heavy duty definition, SRM has its own host group. However, by default, the SRM host group is not provisioned. When creating a cluster with the heavy duty definition, you must set the instance count of the Srm nodes host group to at least one. Otherwise, SRM is not be deployed on the cluster.

For more information on cluster provisioning, see *Creating your first Streams Messaging cluster*. For more information on the default cluster definitions and cluster layouts, see *Streams Messaging cluster layout*.



Note: Deploying SRM in a Data Hub cluster requires version 7.2.6 or higher of Cloudera Runtime.

Prerequisites for using SRM

SRM can be used to replicate Kafka data between all types of CDP clusters. However, the following conditions must be met for all deployments and use cases:

- SRM must be able to access the Kafka hosts of the source and target cluster through the network.
- SRM must trust the TLS certificates of the brokers in the source and target clusters.

This is required so that SRM can establish a trusted connection.

- SRM must have access to credentials that it can use to authenticate itself in both the source and target clusters.
- SRM must use a principal that is authorized to access Kafka resources (topics) on both source and target clusters.

Cloud-based use cases for SRM

There are three common uses cases when using SRM in a cloud-based context. These are as follows:

- Replicating data from a CDP PvC Base cluster to a Data Hub cluster with SRM running in the CDP PvC Base cluster.



Note: Although in this use case SRM is not deployed in a Data Hub cluster, it is still considered as a cloud based use case as either the source or the destination of the replicated data is a Data Hub cluster.

- Replicating data from a CDP PvC Base cluster to a Data Hub cluster with SRM running in the Data Hub cluster.
- Replicating data between two Data Hub clusters.

Related Information

[Creating your first Streams Messaging cluster](#)

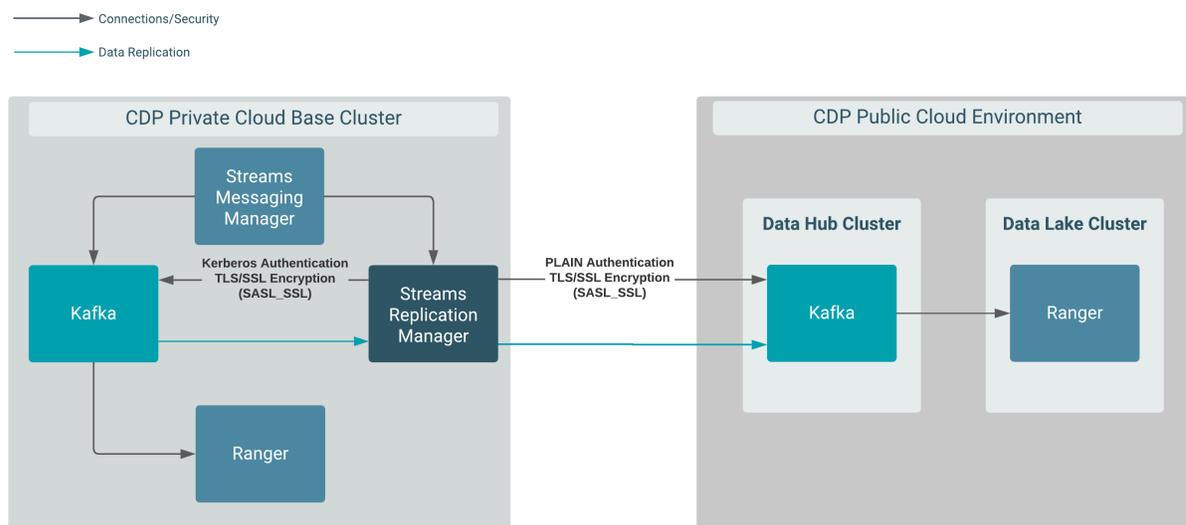
[Streams Messaging cluster layout](#)

Replicating data from CDP PvC Base cluster to Data Hub cluster with SRM running in CDP PvC Base cluster

You can set up and configure an instance of SRM running in a CDP PvC Base cluster to replicate data between the CDP PvC Base cluster and a Data Hub cluster. In addition, you can use SMM to monitor the replication process. Review the following example to learn how this can be set up.

About this task

Consider the following replication scenario:



In this scenario, data is replicated from a CDP PvC Base cluster that has Kafka, SRM, and SMM deployed on it. This is a secure cluster that has TLS/SSL encryption and Kerberos authentication enabled. In addition, it uses Ranger for authorization.

Data is being replicated from this cluster by SRM deployed in this cluster to a Data Hub cluster.

The Data Hub cluster is provisioned with one of the default Streams Messaging cluster definitions.

Before you begin

This example scenario does not go into detail on how to set up the clusters and assumes the following:

- A Data Hub cluster provisioned with the Streams Messaging Light Duty or Heavy Duty cluster definition is available.

For more information, see [Creating your first Streams Messaging cluster](#) in the CDF for Data Hub library. Alternatively, you can also review the cloud provider specific cluster creation instructions available in the [Cloudera Data Hub library](#).

- A CDP PvC Base cluster with Kafka, SRM, and SMM is available. This cluster is TLS/SSL and Kerberos enabled. In addition, it uses Ranger for authorization.

For more information, see the [CDP Private Cloud Base Installation Guide](#).

- Network connectivity and DNS resolution are established between the clusters.

This example scenario demonstrates the configuration required to enable replication monitoring of the Data Hub cluster with Streams Messaging Manager. This can be done by configuring the SRM Service role to target (monitor) the Data Hub cluster. This is done as the last step in the following list of steps and is marked optional. This is because enabling replication monitoring of the Data Hub cluster results in a number of caveats, which are the following:

- The SRM Service role will generate additional cloud traffic.
Any extra traffic you might have in your cloud deployment can lead to additional cloud costs.
- The replications tab in SMM will display all replications targeting the Data Hub cluster.

Although this is expected, you must understand that all other pages in SMM will display information regarding the CDP PvC cluster. A setup like this might lead to confusion or mislead users on what this specific instance of SMM is monitoring.

- You will lose the ability to monitor the replications targeting the CDP PvC cluster.

This is only critical if you have any existing replications that are targeting the CDP PvC cluster and you are monitoring these replications with the SMM instance running in the CDP PvC cluster.



Important: In the following scenario, a new CDP machine user is created and set up specifically for SRM. Alternatively, it is also possible to use an existing machine user and skip steps 1 through 3, but this can only be done if the following requirements are met:

- The existing machine user has access to your CDP environment.
- The existing machine user has the correct Ranger permissions assigned to it.
- You have access to the existing machine user's credentials.

Procedure

1. Create a machine user for SRM in Management Console:

A machine user is required so that SRM has credentials that it can use to connect to the Kafka service in the Data Hub cluster.

- a) Navigate to Management Console User Management.
- b) Click Actions Create Machine User.
- c) Enter a unique name for the user and click Create.

For example: srm

After the user is created, you are presented with a page that displays the user details.



Note:

The Workload User Name (srv_srm), is different from the actual machine user name (srm). The Workload User Name is the identifier you use to configure SRM.

- d) Click Set Workload Password.
- e) Type a password in the Password and Confirm Password fields. Leave the Environment field blank.
- f) Click Set Workload Password.
A message appears on successful password creation.

2. Grant the machine user access to your environment:

You must grant the machine user access to your environment for SRM to connect to the Kafka service with this user.

a) Navigate to Management Console Environments, and select the environment where your Kafka cluster is located.

b) Click Actions Manage Access.

Use the search box to find and select the machine user you want to use.

A list of **Resource Roles** appears.

c) Select the EnvironmentUser role and click Update Roles.

d) Go back to the **Environment Details** page and click Actions Synchronize Users to FreeIPA .

e) On the **Synchronize Users** page, click Synchronize Users.

Synchronizing users ensures that the role assignment is in effect for the environment.



Important: Wait until this process is completed. Otherwise, you will not be able to continue with the next step.

3. Add Ranger permissions for the user you created for SRM in the Data Hub cluster:

You must to grant the necessary privileges to the user so that the user can access Kafka resources. This is configured through Ranger policies.

a) Navigate to Management Console Environments, and select the environment where your Kafka cluster is located.

b) Click the Ranger link on the Environment Details page.

c) Select the resource-based service corresponding to the Kafka resource in the Data Hub cluster.

d) Add the Workload User Name of the user you created for SRM to the following Ranger policies:

- All - consumergroup
- All - topic
- All - transactionalid
- All - cluster
- All - delegationtoken

4. Ensure that Ranger permissions exist for the streamsrepmgr user in the CDP PvC Base cluster:

a) Access the Cloudera Manager instance of your CDP PvC Base cluster.

b) Go to Ranger Ranger Admin Web UI.

c) Log in to the Ranger Console (Ranger Admin Web UI).

d) Ensure that the streamsrepmgr user is added to all required policies.

If the user is missing, add it. The required policies are as follows:

- All - consumergroup
- All - topic
- All - transactionalid
- All - cluster
- All - delegationtoken

5. Create a truststore on the CDP PvC Base cluster:

A truststore is required so that the SRM instance running in the CDP PvC Base cluster can trust the secure Data Hub cluster. To do this, you extract the FreeIPA certificate from the CDP environment, create a truststore that includes the certificate, and copy the truststore to all hosts on the CDP PvC Base cluster.

- a) Navigate to Management Console Environments, and select the environment where your Kafka cluster is located.
- b) Go to the Summary tab.
- c) Scroll down to the FreeIPA section.
- d) Click ActionsGet FreeIPA Certificate.
The FreeIPA certificate file, `[***ENVIRONMENT NAME***]-env.crt`, is downloaded to your computer.
- e) Run the following command to create the truststore:

```
keytool \
  -importcert \
  -storetype JKS \
  -noprompt \
  -keystore datahub-truststore.jks \
  -storepass [***PASSWORD***] \
  -alias freeipa-ca \
  -file [***PATH TO FREEIPA CERTIFICATE***]
```

- f) Copy the `datahub-truststore.jks` file to a common location on all the hosts in your CDP PvC Base cluster. Cloudera recommends that you use the following location: `/opt/cloudera/security/datahub-truststore.jks`.
- g) Set the correct file permissions.
Use 751 for the directory and 444 for the truststore file.

6. Access the Cloudera Manager instance of your CDP PvC Base cluster.**7. Define the external Kafka cluster (Data Hub):**

- a) Go to AdministrationExternal Accounts.
- b) Go to the Kafka Credentials tab.
On this tab you will create a credential for each external cluster taking part in the replication process.
- c) Click Add Kafka credentials.
- d) Configure the Kafka credentials:

In the case of this example, you must create a single credential representing the Data Hub cluster. For example:

```
Name=datahub
Bootstrap servers=[**MY-DATAHUB-CLUSTER-HOST-1.COM:9093**],[**MY-DATAHUB-CLUSTER-HOST-1.COM:9093**]
Security Protocol=SASL_SSL
JAAS Secret 1=[**WORKLOAD USER NAME**]
JAAS Secret 2=[**MACHINE USER PASSWORD**]
JAAS Template=org.apache.kafka.common.security.plain.PlainLoginModule r
required username="##JAAS_SECRET_1##" password="##JAAS_SECRET_2##";
SASL Mechanism=PLAIN
Truststore Password=[**PASSWORD**]
Truststore Path=/opt/cloudera/security/datahub-truststore.jks
Truststore type=JKS
```

- e) Click Add.

If credential creation is successful, a new entry corresponding to the Kafka credential you specified appears on the page.

8. Define the co-located Kafka cluster (PvC Base):

- a) In Cloudera Manager, go to Clusters and select the Streams Replication Manager service.
- b) Go to Configuration.
- c) Find and enable the Kafka Service property.
- d) Find and configure the Streams Replication Manager Co-located Kafka Cluster Alias property.

The alias you configure represents the co-located cluster. Enter an alias that is unique and easily identifiable. For example:

```
cdppvc
```

- e) Enable relevant security feature toggles.

Because CDP PvC Base is both TLS/SSL and Kerberos enabled, you must enable all feature toggles for both the Driver and Service roles. The feature toggles are the following:

- Enable TLS/SSL for SRM Driver
- Enable TLS/SSL for SRM Service
- Enable Kerberos Authentication

9. Add both clusters to SRM's configuration:

- a) Find and configure the External Kafka Accounts property.

Add the name of all Kafka credentials you created to this property. This can be done by clicking the add button to add a new line to the property and then entering the name of the Kafka credential. For example:

```
datahub
```

- b) Find and configure the Streams Replication Manager Cluster alias property.

Add all cluster aliases to this property. This includes the aliases present in both the External Kafka Accounts and Streams Replication Manager Co-located Kafka Cluster Alias properties. Delimit the aliases with commas. For example:

```
datahub,cdppvc
```

10. Configure replications:

In this example data is replicated unidirectionally. As a result, only a single replication must be configured.

- a) Find the Streams Replication Manager's Replication Configs property.
- b) Click the add button and add new lines for each unique replication you want to add and enable.
- c) Add and enable your replications. For example:

```
cdppvc->datahub.enabled=true
```

11. Configure Driver and Service role targets:

- a) Find and configure the Streams Replication Manager Service Target Cluster property.

Add the co-located cluster's alias to the property. For example:

```
cdppvc
```

Setting this property to `cdppvc` does not enable you to monitor the replications targeting the Data Hub cluster. It is possible to add the Data Hub cluster alias to this property and as a result monitor the Data Hub cluster. However, this can lead to unwanted behaviour. See the *Before you begin* section for more information.

- b) Find and configure the Streams Replication Manager Driver Target Cluster property.

For example:

```
datahub, cdppvc
```



Note: This property must either contain all aliases or left blank. Leaving the property blank has the same effect as adding all aliases.

12. Configure the srm-control tool:

- a) Click Gateway in the **Filters** pane.

- b) Find and configure the following properties:

- SRM Client's Secure Storage Type: PKCS12
- SRM Client's Secure Storage Password: [***PASSWORD***]
- Environment Variable Holding SRM Client's Secure Storage Password: SECURESTOREPASS
- Gateway TLS/SSL Trust Store File: [***CDP PVC BASE GLOBAL TRUSTSTORE LOCATION***]
- Gateway TLS/SSL Truststore Password: [***CDP PVC BASE GLOBAL TRUSTSTORE PASSWORD***]
- SRM Client's Kerberos Principal Name: [***MY KERBEROS PRINCIPAL***]
- SRM Client's Kerberos Keytab Location: [***PATH TO KEYTAB FILE***]

Take note of the password you configure in SRM Client's Secure Storage Password and the name you configure in Environment Variable Holding SRM Client's Secure Storage Password. You will need to provide both of these in your CLI session before running the tool.

- c) Click Save Changes.
d) Restart the SRM service.
e) Deploy client configuration for SRM.

13. Start the replication process using the srm-control tool:

- a) SSH as an administrator to any of the SRM hosts in the CDP PvC cluster.

```
ssh [***USER***]@[***MY-CDP-PVC-CLUSTER.COM***]
```

- b) Set the secure storage password as an environment variable.

```
export [***SECURE STORAGE ENV VAR***]="[***SECURE STORAGE PASSWORD***]"
```

Replace [***SECURE STORAGE ENV VAR***] with the name of the environment variable you specified in Environment Variable Holding SRM Client's Secure Storage Password. Replace [***SRM SECURE

`STORAGE PASSWORD***]` with the password you specified in SRM Client's Secure Storage Password. For example:

```
export SECURESTOREPASS="mypassword"
```

- c) Use the `srm-control` tool with the `topics` subcommand to add topics to the allow list.

```
srm-control topics --source cdppvc --target datahub --add [***TOPIC  
NAME***]
```

- d) Use the `srm-control` tool with the `groups` subcommand to add groups to the allow list.

```
srm-control groups --source cdppvc --target datahub --add ".*"
```

14. Configure replication monitoring of the Data Hub cluster:

- Access the Cloudera Manager instance of your CDP PvC Base cluster.
- In Cloudera Manager, go to Clusters and select the Streams Replication Manager service.
- Go to Configuration.
- Find and configure the Streams Replication Manager Service Target Cluster property.

Replace the alias set in the property with the Data Hub cluster's alias. For example:

```
datahub
```

- Click Save Changes.
- Restart the SRM service.
- Access the SMM UI in the CDP PvC Base cluster and go to the Cluster Replications page.

The replications you set up will be visible on this page.



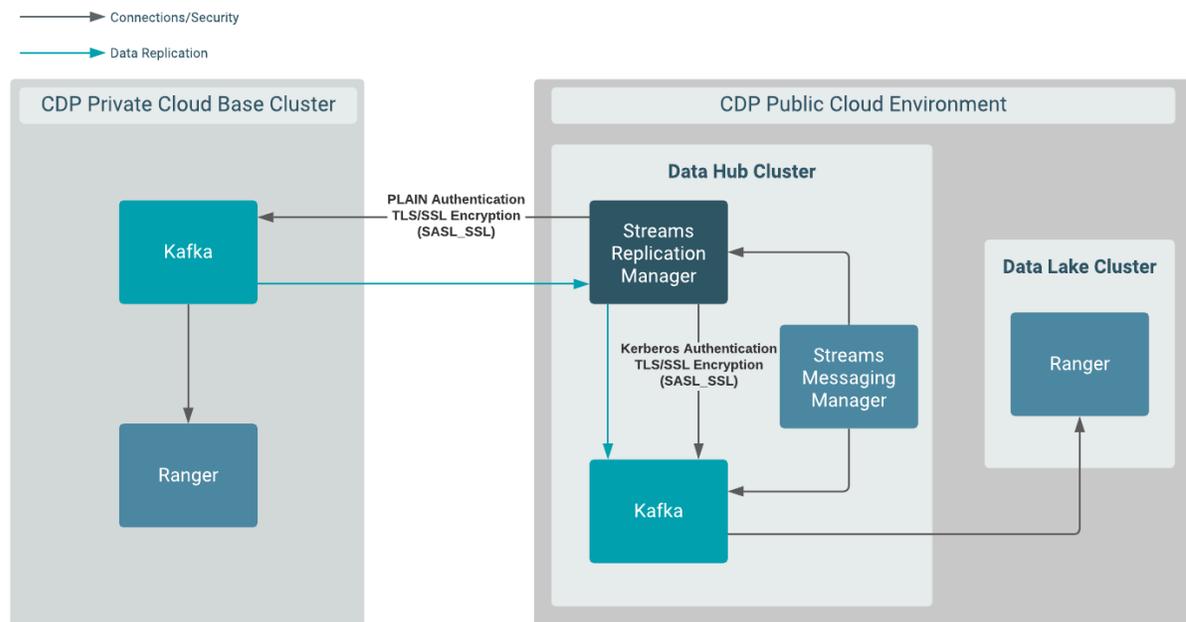
Note: If the topics or groups you added for replication are newly created, they might not be immediately visible. This is due to how frequently SRM checks for newly created topics and consumer groups. By default, this is set to 10 minutes, but can be configured with the Refresh Topics Interval Seconds and Refresh Groups Interval Seconds SRM properties. If at first your topics do not appear, wait a few minutes and refresh the page.

Replicating data from CDP PvC Base cluster to Data Hub cluster with SRM deployed in Data Hub cluster

You can set up and configure an instance of SRM running in a Data Hub cluster to replicate data between the Data Hub cluster and a CDP PvC Base cluster. In addition, you can use SMM to monitor the replication process. Review the following example to learn how this can be set up.

About this task

Consider the following replication scenario:



In this scenario, data is replicated from a CDP PvC Base cluster to a Data Hub cluster by an SRM instance that is deployed in the Data Hub cluster.

The CDP PvC Base cluster has Kafka deployed on it. It is a secure cluster that has TLS/SSL encryption enabled and uses PLAIN authentication. In addition, it uses Ranger for authorization.

The Data Hub cluster is provisioned with the one of the default Streams Messaging cluster definitions.

Before you begin

This example scenario does not go into detail on how to set up the clusters and assumes the following:

- A Data Hub cluster provisioned with the Streams Messaging Light Duty or Heavy Duty cluster definition is available.

For more information, see [Creating your first Streams Messaging cluster](#) in the CDF for Data Hub library. Alternatively, you can also review the cloud provider specific cluster creation instructions available in the [Cloudera Data Hub library](#).

- A CDP PvC Base cluster with Kafka is available. This cluster has TLS/SSL encryption enabled, uses PLAIN authentication, and has Ranger for authorization. For more information, see the [CDP Private Cloud Base Installation Guide](#).
- Network connectivity and DNS resolution are established between the clusters.

Procedure

1. Obtain PLAIN credentials for SRM.

The credentials of a PLAIN user that can access the CDP PvC Base cluster are required. These credentials are supplied to SRM in a later step. In this example `[***PLAIN USER***]` and `[***PLAIN USER PASSWORD***]` is used to refer to these credentials.



Note: Typically, Kerberos cannot be easily configured to span across an on-premise and a public cloud environment. Therefore, these instructions assume that these environments use authentication methods that are easy to interoperate in such hybrid environments like LDAP or PAM authentication. For more information on how you can configure your Kafka service to use LDAP or PAM, see [Kafka Authentication](#).

2. Add Ranger permissions for the PLAIN user in the CDP PvC cluster:

You must ensure that the PLAIN user you obtained has correct permissions assigned to it in Ranger. Otherwise, SRM will not be able to access Kafka resources on the CDP PvC Base cluster.

- a) Access the Cloudera Manager instance of your CDP PvC Base cluster.
- b) Go to Ranger Ranger Admin Web UI.
- c) Log in to the Ranger Console (Ranger Admin Web UI).
- d) Add the [***PLAIN USER***] to the following policies:
 - All - consumergroup
 - All - topic
 - All - transactionalid
 - All - cluster
 - All - delegationtoken

3. Acquire the CDP PvC Base cluster truststore and add it to the Data Hub cluster:

The actions you need to take differ depending on how TLS is set up in the CDP PvC Base cluster:

For Auto TLS**a. Obtain the certificate of the Cloudera Manager root Certificate Authority and its password.**

The Certificate Authority certificate and its password can be obtained using the Cloudera Manager API. The following steps describe how you can retrieve the certificate and password using the Cloudera Manager API Explorer. Alternatively, you can also retrieve the certificate and password by calling the appropriate endpoints in your browser window or using curl.

1. Access the Cloudera Manager instance of your CDP PvC Base cluster.
2. Go to SupportAPI Explorer.
3. Find CertManagerResource.
4. Select the /certs/truststore GET operation and click Try it out.
5. Enter the truststore type.
6. Click Execute.
7. Click Download file under Responses.

The downloaded file is your certificate.

8. Select the /certs/truststorePassword GET operation and click Try it out.
9. Click Execute.

The password is displayed under Responses.

b. Run the following command to create the truststore:

```
keytool \  
-importcert \  
-storetype JKS \  
-noprompt \  
-keystore cdppvc-truststore.jks \  
-storepass ***PASSWORD*** \  
-alias cdppvc-cm-ca \  

```

```
-file ***PATH TO CM CA CERTIFICATE***
```

Note down the password, it is needed in a later step.

- c. Copy the `cdpdc-truststore.jks` file to a common location on all the hosts in your CDP Data Hub cluster.
Cloudera recommends that you use the following location: `/opt/cloudera/security/cdppvc-truststore.jks`.
- d. Set the correct file permissions.
Use 751 for the directory and 444 for the truststore file.

For Manual TLS

- a. Note down the CDP PvC Base cluster's truststore location and password, these should be known to you.
- b. Copy the truststore file to a common location on all the hosts in your CDP Data Hub cluster.
Cloudera recommends that you use the following location: `/opt/cloudera/security/truststore.jks`.
- c. Set the correct file permissions.
Use 751 for the directory and 444 for the truststore file.

- 4. Access the Cloudera Manager instance of your Data Hub Cluster.
- 5. Define the external Kafka cluster (CDP PvC Base).

- a) Go to Administration External Accounts.
- b) Go to the Kafka Credentials tab.

On this tab you will create a credential for each external cluster taking part in the replication process.

- c) Click Add Kafka credentials
- d) Configure the Kafka credentials:

In the case of this example, you must create a single credential representing the CDP PvC Base cluster. For example:

```
Name=cdppvc
Bootstrap servers=[***MY-CDP-PVC-CLUSTER-HOST-1.COM:9093***], [***MY-CDP-PVC-CLUSTER-HOST-2:9093***]
Security Protocol=SASL_SSL
JAAS Secret 1=[***PLAIN USER***]
JAAS Secret 2=[***PLAIN USER PASSWORD***]
JAAS Template=org.apache.kafka.common.security.plain.PlainLoginModule r
equired username="##JAAS_SECRET_1##" password="##JAAS_SECRET_2##";
SASL Mechanism=PLAIN
Truststore Password=[***PASSWORD***]
Truststore Path=/opt/cloudera/security/cdppvc-truststore.jks
Truststore type=JKS
```



Note: The properties you specify for the Kafka credential depend on the security configuration of the CDP PvC Base cluster. This specific example is for a cluster that has TLS/SSL encryption and PLAIN authentication enabled. You must change these configurations based on the setup of your CDP PvC Base cluster.

- e) Click Add.

If credential creation is successful, a new entry corresponding to the Kafka credential you specified appears on the page.

6. Define the co-located Kafka cluster (Datahub):

Note: Some of the following properties might already be configured by automation.

- a) In Cloudera Manager, go to Clusters and select the Streams Replication Manager service.
- b) Go to Configuration.
- c) Find and enable the Kafka Service property.
- d) Find and configure the Streams Replication Manager Co-located Kafka Cluster Alias property.

The alias you configure represents the co-located cluster. Enter an alias that is unique and easily identifiable. For example:

```
datahub
```

- e) Enable relevant security feature toggles.

Because the Data Hub cluster is both TLS/SSL and Kerberos enabled, you must enable all feature toggles for both the Driver and Service roles. The feature toggles are the following:

- Enable TLS/SSL for SRM Driver
- Enable TLS/SSL for SRM Service
- Enable Kerberos Authentication

7. Add both clusters to SRM's configuration:

- a) Find and configure the External Kafka Accounts property.

Add the name of all Kafka credentials you created to this property. This can be done by clicking the add button to add a new line to the property and then entering the name of the Kafka credential. For example:

```
cdppvc
```

- b) Find and configure the Streams Replication Manager Cluster alias property.

Add all cluster aliases to this property. This includes the aliases present in both the External Kafka Accounts and Streams Replication Manager Co-located Kafka Cluster Alias properties. Delimit the aliases with commas. For example:

```
datahub,cdppvc
```

8. Configure replications:

In this example data is replicated unidirectionally. As a result, only a single replication must be configured.

- a) Find the Streams Replication Manager's Replication Configs property.
- b) Click the add button and add new lines for each unique replication you want to add and enable.
- c) Add and enable your replications. For example:

```
cdppvc->datahub.enabled=true
```

9. Configure Driver and Service role targets:

- a) Find and configure the Streams Replication Manager Service Target Cluster property.
Add the co-located cluster's alias to the property. For example:

```
datahub
```

- b) Find and configure the Streams Replication Manager Driver Target Cluster property.
For example:

```
datahub,cdppvc
```



Important: If you have another SRM instance configured with the same clusters and is targeting the CDP PvC Base cluster, the `cdppvc` alias should not be configured as a target for this instance of SRM.



Note: This property must either contain all aliases or left blank. Leaving the property blank has the same effect as adding all aliases.

10. Configure the `srm-control` tool:

- a) Click Gateway in the **Filters** pane.
- b) Find and configure the following properties:
- SRM Client's Secure Storage Type: PKCS12
 - SRM Client's Secure Storage Password: [***PASSWORD***]
 - Environment Variable Holding SRM Client's Secure Storage Password: SECURESTOREPASS
 - Gateway TLS/SSL Trust Store File: `/opt/cloudera/security/datahub-truststore.jks`
 - Gateway TLS/SSL Truststore Password: [***PASSWORD***]
 - SRM Client's Kerberos Principal Name: [***MY KERBEROS PRINCIPAL***]
 - SRM Client's Kerberos Keytab Location: [***PATH TO KEYTAB FILE***]

Take note of the password you configure in SRM Client's Secure Storage Password and the name you configure in Environment Variable Holding SRM Client's Secure Storage Password. You will need to provide both of these in your CLI session before running the tool.

- c) Click Save Changes.
- d) Restart the SRM service.
- e) Deploy client configuration for SRM.

11. Start the replication process using the `srm-control` tool:

- a) SSH as an administrator to any of the SRM hosts in the Data Hub cluster.

```
ssh [***USER***]@[***MY-DATAHUB-CLUSTER.COM***]
```

- b) Set the secure storage password as an environment variable.

```
export [***SECURE STORAGE ENV VAR***]="[***SECURE STORAGE PASSWORD***]"
```

Replace [***SECURE STORAGE ENV VAR***] with the name of the environment variable you specified in Environment Variable Holding SRM Client's Secure Storage Password. Replace [***SRM SECURE

`STORAGE_PASSWORD***]` with the password you specified in SRM Client's Secure Storage Password. For example:

```
export SECURESTOREPASS="mypassword"
```

- c) Use the `srm-control` tool with the `topics` subcommand to add topics to the allow list.

```
srm-control topics --source cdppvc --target datahub --add [***TOPIC
NAME***]
```

- d) Use the `srm-control` tool with the `groups` subcommand to add groups to the allow list.

```
srm-control groups --source cdppvc --target datahub --add ".*"
```

12. Monitor the replication process.

Access the SMM UI in the Data Hub cluster and go to the Cluster Replications page. The replications you set up will be visible on this page.



Note: If the topics or groups you added for replication are newly created, they might not be immediately visible. This is due to how frequently SRM checks for newly created topics and consumer groups. By default, this is set to 10 minutes, but can be configured with the Refresh Topics Interval Seconds and Refresh Groups Interval Seconds SRM properties. If at first your topics do not appear, wait a few minutes and refresh the page.

Related Information

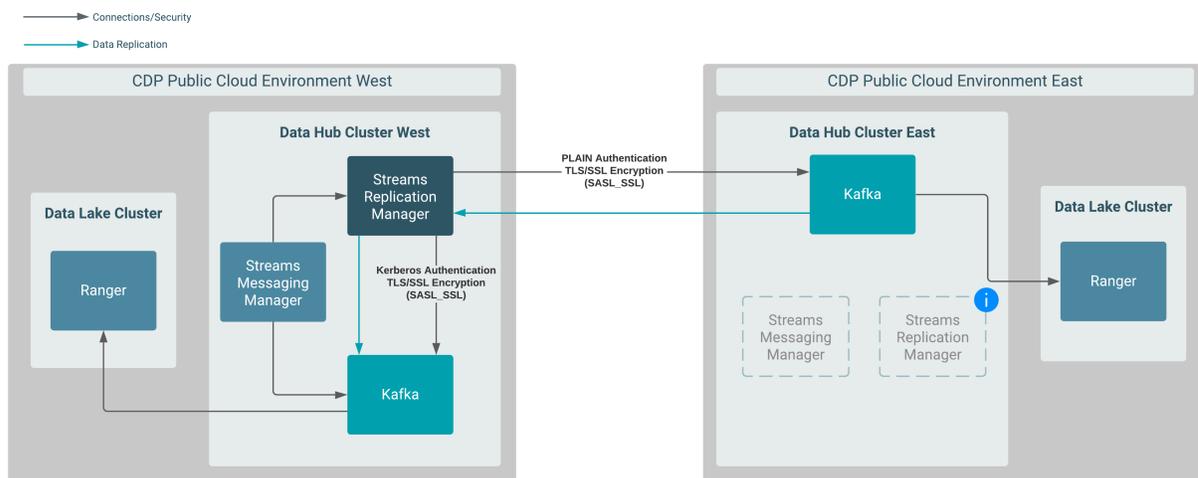
[Kafka Authentication](#)

Replicating data between Data Hub clusters with SRM deployed in a Data Hub cluster.

You can set up and configure an instance of SRM in a Data Hub cluster to replicate data between Data Hub clusters. In addition, you can use SMM to monitor the replication process. Review the following example to learn how this can be set up.

About this task

Consider the following replication scenario:



In this scenario, data is replicated between two Data Hub clusters that are provisioned in different CDP environments. More specifically, data in Data Hub East is replicated to Data Hub West by an instance of SRM running in Data Hub West.

Both Data Hub clusters are provisioned with the default Streams Messaging cluster definitions.

SRM and SMM are available in both clusters, but the instances in Data Hub East are not utilized in this scenario.

Before you begin

This example scenario does not go into detail on how to set up the clusters and assumes the following:

- Two Data Hub clusters provisioned with the Streams Messaging Light Duty or Heavy Duty cluster definition are available.

For more information, see [Creating your first Streams Messaging cluster](#) in the CDF for Data Hub library. Alternatively, you can also review the cloud provider specific cluster creation instructions available in the [Cloudera Data Hub library](#).

- Network connectivity and DNS resolution are established between the clusters.



Important: In the following scenario, a new CDP machine user is created and set up specifically for SRM. Alternatively, it is also possible to use an existing machine user and skip steps 1 through 3, but this can only be done if the following requirements are met:

- The existing machine user has access to your CDP environment.
- The existing machine user has the correct Ranger permissions assigned to it.
- You have access to the existing machine user's credentials.

Procedure

1. Create a machine user for SRM in Management Console:

A machine user is required so that SRM has credentials that it can use to connect to the Kafka service in the Data Hub cluster. This step is only required in the environment where SRM is not running. In the case of this example, this is the CDP Public Cloud East environment.

- a) Navigate to Management Console > User Management.
- b) Click Actions > Create Machine User.
- c) Enter a unique name for the user and click Create.

For example: srm

After the user is created, you are presented with a page that displays the user details.



Note:

The Workload User Name (srv_srm), is different from the actual machine user name (srm). The Workload User Name is the identifier you use to configure SRM.

- d) Click Set Workload Password.
- e) Type a password in the Password and Confirm Password fields. Leave the Environment field blank.
- f) Click Set Workload Password.

A message appears on successful password creation.

2. Grant the machine user access to your environment:

You must to grant the machine user access in your environments, otherwise SRM will not be able to connect to the Kafka service with this user. This step is only required in the environments where SRM is not running. In the case of this example this is the CDP Public Cloud East environment.

- a) Navigate to Management Console Environments, and select the environment where your Kafka cluster is located.
- b) Click Actions Manage Access.
Use the search box to find and select the machine user you want to use.
A list of **Resource Roles** appears.
- c) Select the EnvironmentUser role and click Update Roles.
- d) Go back to the **Environment Details** page and click Actions Synchronize Users to FreeIPA .
- e) On the **Synchronize Users** page, click Synchronize Users.

Synchronizing users ensures that the role assignment is in effect for the environment.



Important: Wait until this process is completed. Otherwise, you will not be able to continue with the next step.

3. Add Ranger permissions for the user you created for SRM.

This step is only required in the environment where SRM is not running. In the case of this example the environment is the CDP Public Cloud East .

- a) Navigate to Management Console Environments, and select the environment where your Kafka cluster is located.
- b) Click the Ranger link on the Environment Details page.
- c) Select the resource-based service corresponding to the Kafka resource in the Data Hub cluster.
- d) Add the Workload User Name of the user you created for SRM to the following Ranger policies:
 - All - consumergroup
 - All - topic
 - All - transactionalid
 - All - cluster
 - All - delegationtoken

4. Establish trust between the clusters:

A truststore is needed so that the SRM instance running in Data Hub West can trust Data Hub East. To do this, you extract the FreeIPA certificate from Environment East, create a truststore that includes the certificate, and copy the truststore to all hosts on Data Hub West.

- a) Navigate to Management Console Environments, and select Environment East.
- b) Go to the Summary tab.
- c) Scroll down to the FreeIPA section.
- d) Click ActionsGet FreeIPA Certificate.
The FreeIPA certificate file, `[***ENVIRONMENT NAME***]-env.crt`, is downloaded to your computer.
- e) Run the following command to create the truststore:

```
keytool \
  -importcert \
  -storetype JKS \
  -noprompt \
  -keystore truststore-east.jks \
  -storepass [***PASSWORD***] \
  -alias freeipa-east-ca \
  -file [***PATH TO FREEIPA CERTIFICATE***]
```

- f) Copy the truststore-east.jks file to a common location on all the hosts in your Data Hub West cluster.
Cloudera recommends that you use the following location: `/opt/cloudera/security/truststore-east.jks`.

- g) Set the correct file permissions.

Use 751 for the directory and 444 for the truststore file.

5. Access the Cloudera Manager instance of the Data Hub West cluster.

6. Define the external Kafka cluster (Data Hub East):

- a) Go to Administration External Accounts.
b) Go to the Kafka Credentials tab.

On this tab you will create a credential for each external cluster taking part in the replication process.

- c) Click Add Kafka credentials

- d) Configure the Kafka credentials:

In the case of this example, you must create a single credential representing the Data Hub East cluster. For example:

```
Name=dheast
Bootstrap servers=[***MY-DATAHUB-EAST-CLUSTER-
HOST-1.COM:9093***],[***MY-DATAHUB-EAST-CLUSTER-HOST-2:9093***]
Security Protocol=SASL_SSL
JAAS Secret 1=[***WORKLOAD USER NAME***]
JAAS Secret 2=[***MACHINE USER PASSWORD***]
JAAS Template=org.apache.kafka.common.security.plain.PlainLoginModule r
equired username="##JAAS_SECRET_1##" password="##JAAS_SECRET_2##";
SASL Mechanism=PLAIN
Truststore Password=[***PASSWORD***]
Truststore Path=/opt/cloudera/security/truststore-east.jks
Truststore type=JKS
```

- e) Click Add.

If credential creation is successful, a new entry corresponding to the Kafka credential you specified appears on the page.

7. Define the co-located Kafka cluster (Data Hub West):



Note: Some of the following properties might already be configured by automation.

- a) In Cloudera Manager, go to Clusters and select the Streams Replication Manager service.
b) Go to Configuration.
c) Find and enable the Kafka Service property.
d) Find and configure the Streams Replication Manager Co-located Kafka Cluster Alias property.

The alias you configure represents the co-located cluster. Enter an alias that is unique and easily identifiable. For example:

```
dhwest
```

- e) Enable relevant security feature toggles.

Because the Data Hub cluster is both TLS/SSL and Kerberos enabled, you must enable all feature toggles for both the Driver and Service roles. The feature toggles are the following:

- Enable TLS/SSL for SRM Driver
- Enable TLS/SSL for SRM Service
- Enable Kerberos Authentication

8. Add both clusters to SRM's configuration:

- a) Find and configure the External Kafka Accounts property.

Add the name of all Kafka credentials you created to this property. This can be done by clicking the add button to add a new line to the property and then entering the name of the Kafka credential. For example:

```
dheast
```

- b) Find and configure the Streams Replication Manager Cluster alias property.

Add all cluster aliases to this property. This includes the aliases present in both the External Kafka Accounts and Streams Replication Manager Co-located Kafka Cluster Alias properties. Delimit the aliases with commas. For example:

```
dheast , dhwest
```

9. Configure replications:

In this example data is replicated unidirectionally. As a result, only a single replication must be configured.

- Find the Streams Replication Manager's Replication Configs property.
- Click the add button and add new lines for each unique replication you want to add and enable.
- Add and enable your replications. For example:

```
dheast->dhwest.enabled=true
```

10. Configure Driver and Service role targets:

- a) Find and configure the Streams Replication Manager Service Target Cluster property.

Add the co-located cluster's alias to the property. For example:

```
dhwest
```

- b) Find and configure the Streams Replication Manager Driver Target Cluster property.

For example:

```
dheast , dhwest
```



Important: If you have another SRM instance configured with the same clusters and is targeting Data Hub East, the dheast alias should not be configured as a target for this instance of SRM.



Note: This property must either contain all aliases or left blank. Leaving the property blank has the same effect as adding all aliases.

11. Configure the srm-control tool:

- a) Click Gateway in the **Filters** pane.
- b) Find and configure the following properties:



Note: Some of the following properties might already be configured by automation.

- SRM Client's Secure Storage Type: PKCS12
- SRM Client's Secure Storage Password: [***PASSWORD***]
- Environment Variable Holding SRM Client's Secure Storage Password: SECURESTOREPASS
- Gateway TLS/SSL Trust Store File: /opt/cloudera/security/truststore-west.jks
- Gateway TLS/SSL Truststore Password: [***PASSWORD***]
- SRM Client's Kerberos Principal Name: [***MY KERBEROS PRINCIPAL***]
- SRM Client's Kerberos Keytab Location: [***PATH TO KEYTAB FILE***]

Take note of the password you configure in SRM Client's Secure Storage Password and the name you configure in Environment Variable Holding SRM Client's Secure Storage Password. You will need to provide both of these in your CLI session before running the tool.

- c) Click Save Changes.
- d) Restart the SRM service.
- e) Deploy client configuration for SRM.

12. Start the replication process using the srm-control tool:

- a) SSH as an administrator to any of the SRM hosts in the Data Hub West cluster.

```
ssh [***USER***]@[***DATA-HUB-WEST-CLUSTER-HOST-1.COM***]
```

- b) Set the secure storage password as an environment variable.

```
export [***SECURE STORAGE ENV VAR***]="[***SECURE STORAGE PASSWORD***]"
```

Replace [***SECURE STORAGE ENV VAR***] with the name of the environment variable you specified in Environment Variable Holding SRM Client's Secure Storage Password. Replace [***SRM SECURE STORAGE PASSWORD***] with the password you specified in SRM Client's Secure Storage Password. For example:

```
export SECURESTOREPASS="mypassword"
```

- c) Use the srm-control tool with the topics subcommand to add topics to the allow list.

```
srm-control topics --source dheast --target dhwest --add [***TOPIC NAME***]
```

- d) Use the srm-control tool with the groups subcommand to add groups to the allow list.

```
srm-control groups --source dheast --target dhwest --add ".*"
```

13. Monitor the replication process.

Access the SMM UI in the Data Hub West cluster and go to the Cluster Replications page. The replications you set up will be visible on this page.



Note: If the topics or groups you added for replication are newly created, they might not be immediately visible. This is due to how frequently SRM checks for newly created topics and consumer groups. By default, this is set to 10 minutes, but can be configured with the Refresh Topics Interval Seconds and Refresh Groups Interval Seconds SRM properties. If at first your topics do not appear, wait a few minutes and refresh the page.