Cloudera DataFlow for Data Hub 7.2.17

Setting up your Streaming Analytics cluster

Date published: 2019-12-16 Date modified: 2023-06-27



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

Before creating your cluster	4
Assigning resource roles	4
Creating IDBroker mapping	4
Setting workload password	5
Creating your cluster	5
After creating your cluster	7
After creating your cluster	7
After creating your cluster Retrieving keytab file Uploading or unlocking your keytab	7
After creating your cluster Retrieving keytab file Uploading or unlocking your keytab Configuring Ranger policies for Flink and SSB	
After creating your cluster Retrieving keytab file Uploading or unlocking your keytab Configuring Ranger policies for Flink and SSB Configuring Kafka policies.	
After creating your cluster Retrieving keytab file Uploading or unlocking your keytab Configuring Ranger policies for Flink and SSB Configuring Kafka policies Configuring Schema Registry policies	
After creating your cluster Retrieving keytab file Uploading or unlocking your keytab Configuring Ranger policies for Flink and SSB Configuring Kafka policies Configuring Schema Registry policies Configuring Hive policies	

Before creating your cluster

Before you start creating your Streaming Analytics Data Hub cluster, you need to ensure that you have set up the environment properly and have all the necessary accesses to use CDP Public Cloud.

- You have CDP login credentials.
- You have an available CDP environment.
- You have a running Data Lake.
- You have a CDP username and the predefined resource role of this user is EnvironmentAdmin.
- Your CDP user is synchronized to the CDP Public Cloud environment.

Important: Ensure that the Runtime version of the Data Lake cluster matches the Runtime version of the Data Hub cluster that you are about to create. If these versions do not match, you may encounter warnings and/or errors.

Related Information

/!\

Getting started as a user AWS environments Azure environments GCP environments Data lakes

Assigning resource roles

As an administrator, you need to give permissions to users or groups to be able to access and perform tasks in your Data Hub environment.

Procedure

- 1. Navigate to Management Console > Environments and select your environment.
- 2. Click Actions > Manage Access.
- 3. Search for a user or group that needs access to the environment.
- 4. Select EnvironmentUser role from the list of Resource Roles.
- Click Update Roles. The Resource Role for the selected user or group will be updated.
- 6. Navigate to Management Console > Environments , and select the environment where you want to create a cluster.
- 7. Click Actions > Synchronize Users to FreeIPA.
- **8.** Click Synchronize Users.



Note: There might be cases where the status of the environment is synchronized with warnings and has failed status. This does not indicate that the synchronization has failed.

Creating IDBroker mapping

As an administrator, you must create IDBroker mapping for a user or group to access cloud storage. As a part of Knox, the IDBroker allows a user to exchange cluster authentication for temporary cloud credentials.

4

About this task

You must create IDBroker mapping for a user or group to have access to the S3 cloud storage. As a part of Knox, the IDBroker allows a user to exchange cluster authentication for temporary cloud credentials. The following roles are created when registering the CDP environment:

- · idbroker-role: granting permissions to IDBroker instances associated with the CDP environment
- · datalake-admin-role: granting access to CDP cloud resources
- · logs-role: granting access to the logs storage location

For using Streaming Analytics in CDP Public Cloud, you must make sure that the users who run Flink jobs are associated with the ARN of the datalake-admin-role as it grants access to the cloud resources required to run the Flink service.

Procedure

- 1. Navigate to Management Console > Environments and select your environment.
- 2. Click Actions > Manage Access.
- 3. Click on the IDBroker Mappings tab.
- 4. Click Edit to add a new user or group and assign roles to have writing access for the cloud storage.
- 5. Search for the user or group you need to map.
- 6. Go to the IAM Summary page where you can find information about your cloud storage account.
- 7. Copy the Role ARN.
- 8. Go back to the IDBroker Mapping interface on the Cloudera Management Console page.
- 9. Paste the Role ARN to your selected user or group.
- **10.** Click Save and Sync.

Setting workload password

As a user, you need to set a workload password for your EnvironmentUser account to be able to access the SQL Stream Builder nodes through SSH connection.

Procedure

- 1. Navigate to Management Console > Environments and select your environment.
- 2. Click Actions > Manage Access.
- **3.** Click Workload Password.
- 4. Give a chosen workload password for your user.
- 5. Confirm the given password by typing it again.
- 6. Click Set Workload Password.

Creating your cluster

When creating your Streaming Analytics cluster, you must choose from the Light and Heavy duty options, and further select the cluster definition that matches your cloud provider for the environment. You also need to pay attention to the cloud storage settings where Flink saves the checkpoints and savepoints.

About this task

After you have met the prerequisites, you are ready to create your Streaming Analytics cluster using a default cluster definition.

Procedure

- 1. Log into the CDP web interface.
- 2. Navigate to Management Console > Environments , and select the environment where you want to create a cluster.
- 3. Click Create Data Hub.

The following page is displayed:

Management Console	Data Hubs / Provision Data Hub
Pr O Dashboard Pr Dashboard Pr Dashboard Pr Ret Data Lakes	Provision Data Hub trovision on demand workload clusters with the combination of applications for various business needs such as enterprise data warehouse management and data science operations. elected Environment with running Data Lake <u> </u>
옷 User Management	Cluster Definition
Of Data Hub Clusters Image: Data Warehouses	Services Select the Cluster Definition option to create your cluster quickly by using one of the prescriptive cluster definitions included by default or one of your previously created custom cluster definitions.
ラ ML Workspaces 曼 Classic Clusters	Cluster Definition* Please select a Cluster Definition
	General Settings
gy choon or any s	Cluster Name* Enter the cluster name Tags You may optionally add tags, which will help you find your cluster-related resources, such as VMs, in your cloud provider account. Add
→ Get Started	Auto Scaling
HelpHelp	Advanced Options
R: Data Lakes X: User Management Image: Comparison of the comp	 dvxder1 dvxder1

- 4. Select Cluster Definition.
- **5.** Select the Streaming Analytics cluster definition from the Cluster Definition drop-down list depending on your operational objectives.

The following template options are available:

- 7.2.17 Streaming Analytics Light Duty for AWS
- 7.2.17 Streaming Analytics Light Duty for Azure
- 7.2.17 Streaming Analytics Light Duty for GCP
- 7.2.17 Streaming Analytics Heavy Duty for AWS
- 7.2.17 Streaming Analytics Heavy Duty for Azure
- 7.2.17 Streaming Analytics Heavy Duty for GCP

For more information on templates, see *Streaming Analytics Data Hub cluster definitions* and *Streaming Analytics cluster layout*.

The list of services is automatically shown below the selected cluster definition name.

6. Provide a cluster name and add tags you might need.

You can define tags that will be applied to your cluster- related resources on your cloud provider account. For more information about tags, see Tags.

7. Optionally, use the Configure Advanced Options section to customize the infrastructure settings.



Note: Ensure that the right cloud storage path is given in Advanced Options > Cloud Storage .

Cloudera recommends saving the checkpoints and savepoints to the S3 cloud storage to make the saved files available throughout all cluster deployments. You can also use HDFS, however Cloudera only recommends this solution for temporary storage of checkpoints and savepoints.

8. Click Provision Cluster.

Results

You are redirected to the Data Hub cluster dashboard, and a new tile representing your cluster appears at the top of the page.

Related Information Create a cluster from a definition on AWS Create a cluster from a definition on Azure Create a cluster from a definition on GCP Tags Streaming Analytics Data Hub cluster definitons Streaming Analytics cluster layout

After creating your cluster

As an EnvironmentAdmin, you need to provide access to users to your environment and to the Streaming Analytics cluster by assigning user roles, adding users to Ranger policies, and creating IDBroker mappings.

About this task

The cluster you have created using the Streaming Analytics cluster definition is kerberized and secured with SSL. Users can access cluster UIs and endpoints through a secure gateway powered by Apache Knox. Before you can use Flink and SQL Stream Builder, you must provide users access to the Streaming Analytics cluster components.

Related Information IDBroker

Retrieving keytab file

As a user, you need to retrieve the keytab file of your profile and upload it to the Streaming SQL Console to be able to run SQL jobs.

Procedure

- 1. Navigate to Management Console > Environments , and select the environment where you have created your cluster.
- 2. Click on your profile name.
- 3. Click Profile.
- **4.** Click Actions > Get Keytab.
- 5. Choose the environment where your Data Hub cluster is running.
- 6. Click Download.
- 7. Save the keytab file in a chosen location.

Uploading or unlocking your keytab

When accessing the Streaming SQL Console for the first time in Data Hub, you must upload and unlock the keytab file corresponding with your profile before you can use SQL Stream Builder (SSB).

7

Procedure

- **1.** Navigate to the Streaming SQL Console.
 - a) Navigate to Management Console > Environments , and select the environment where you have created your cluster.
 - b) Select the Streaming Analytics cluster from the list of Data Hub clusters.
 - c) Select Streaming SQL Console from the list of services.

The Streaming SQL Console opens in a new window.

2. Click your username on the sidebar of the Streaming SQL Console.

3. Click Manage keytab.

The Keytab Manager window appears.

You can either unlock the keytab already existing on the cluster, or you can directly upload your keytab file in the SQL Stream Builder.

a) Unlock your keytab by providing the Principal Name and Password, and clicking Unlock Keytab. The Principal Name and Password should be the same as the workload username and password set for the Streaming Analytics cluster.

Keytab	Keytab Manager					
Unlock	Upload					
Principal	Name *					
Passwor	d *					

Lock Keytab

Cancel

Unlock Keytab

b) Upload your keytab by clicking on the Upload tab, uploading the keytab file directly to the Console, and clicking Unlock Keytab.

Unlock	Upload			
Principal Na	ame *			
Choos	e File	No file choose	en	
Lock Kevta	b		Cance	Upload Ke

- **a.** Retrieve your keytab file.
 - 1. Click on your profile name in the Management Console.
 - **2.** Click Profile.
 - **3.** Click Actions > Get Keytab.
 - 4. Choose the environment where your Data Hub cluster is running.
 - 5. Click Download.
 - 6. Save the keytab file in a chosen location.
- **b.** Manually upload your keytab to the Streaming Analytics cluster:

```
scp <location>/<your_keytab_file> <workload_username>@<manager_node_FQDN
>:.
```

Password:<your_workload_password>

c. Access the manager node of your Streaming Analytics cluster:

```
ssh <workload_username>@<manager_node_FQDN>
```

Password: <workload_password>

d. Use kinit command to authenticate your user:

kinit -kt <keytab_filename>.keytab <workload_username>

e. Use the flink-yarn-session command to see if the authentication works properly:

```
flink-yarn-session -d \
-D security.kerberos.login.keytab=<keytab_filename>.keytab \
-D security.kerberos.login.principal=<workload_username>
```

In case the command fails, you can review the log file for further information about the issue.

Configuring Ranger policies for Flink and SSB

You must add your workload username and the SQL Stream Builder (SSB) service user to the Ranger policies that are needed for Kafka, Schema Registry, Hive and Kudu to provide access to topics, schemas and tables used by the components and to be able to execute Flink jobs.

About this task

You need to provide access to users and the SSB service by configuring Ranger policies for the Kafka data source and the Schema Registry, Kudu and Hive catalog services. To be able to use Flink, you need to add the workload user or users to the required policies. For SSB, the ssb service user needs to be added to the same policies.

When adding more workload users, instead of adding them one by one, you can create user groups in Ranger, for example a flink_users group. This way you can assign every user who will use the Streaming Analytics cluster into a group, and add only that one group to the Ranger policies.

Procedure

1. Navigate to Management Console > Environments, and select the environment where you have created your cluster.

2. Click Ranger from the Quick Links or select Data Lake Ranger from Services.

Hanagement Console	Environments /	/ Data Lake / Event H	listory	
 Dashboard Environments Data Lakes 	Crm.cdp.emvironments.us.wc US West (Oregon)-us.wet	11-oregon 0 \$1-13074cee4-1cad-4507-b645-7ccf9edbb73d environment #2	cr/21865e-2786-42ac-8674-4bbbc1a9e795 🛱	Stop Actions •
유, User Management	sdx Data Lake Details NAME STATUS © Running	NODES 2 @ 0 0 0 STATUS REASON Datalake cert renewal finished	SCALE Light Duty CRN crr:cdp:datalake:us-west-1:9d74eei	QUICK LINKS Atlas Ranger Data Catalog 4-1cad-45d7-b645-7ccf9edbb73d:datalake:8f5d4de0-bcf6-47ac-8c1a-8_ (2)
A Data Services Clusters Audit Sared Resources	Data Hubs Data Lake FreeIPA	Data Services Clusters Cluster Definitions	Summary	
۞ Global Settings	Environment Details	CREDENTIAL REG US-V	ION AVAILABILIT vest-2 N/A	YZONE
 → Get Started ③ Help ④ 	Services Image: Atlas Image: Solr Server	CM-UI	. HBase UI 🥪 Name	Node 🤯 Ranger
2.68.0-b86 «	Cloudera Manager Info			

You are redirected to the Ranger Admin Web user interface (UI) where you can add the workload user and SSB service user to the required policies to grant access for Flink and SSB.

Ŕ	Ranger	C Access Manager	🗋 Audit	Security Zone	Settings			🙀 csso_dmarsal 🝷
	Service Manager						Last Respons	e Time : 02/05/2023 06:07:28 PM
s	ervice Manager	r					Security Zone: Select Zone Name	* 🖸 Import 🖾 Export
	🕞 HDFS			+ 🛛 🖓		+ 2 2	- HADOOP SQL	+ 2 2
	cm_hdfs			• 2 •	cm_hbase	• •	Hadoop SQL	
				+ 🛛 🖓	🕞 кнох	+ 2 2		+ 🛛 🗖
	cm_yam				cm_knox		cm_solr	• • •
	csa_flink_cluster	5_yam		• 2 •				
				+ 0 0		+ 🛛 🖓		+ 🛛 🖸
	cm_kafka			• 2 •	cfm_nifi_cluster_1_nifi_NIFI_BASE	• •	cfm_nifi_cluster_1_nifiregistry	
	csa_flink_cluster	5_kafka_0d8c		• 7				
	csm_kafka_clust	er_1_kafka_d478		• 7 •				
				+ 🛛 🖸	ADLS	+ 🛛 🖸	E KUDU	+ 🛛 🖸
	cm_atlas			• 2 0	cm_adls	• 7 1	cm_kudu	• • •
				+ 🛛 🖸		+ 🛛 🖸		+ 🛛 🖸
	cm_ozone			• 7	csm_kafka_cluster_1_schemaregistry	• 2 •	cm_kafka_connect	• 7
							csm_kafka_cluster_1_kafka_connect	

Configuring Kafka policies

After accessing the Ranger Admin Web UI, the workload username or user groups, and the SSB service user needs to be added to the Kafka policies to be able to use Flink and SSB.

About this task

The following resource based policies need to be configured for the Kafka service:

- all topic: Provides access to all topics for users
- all consumergroup: Provides access to all consumer groups for users
- all cluster: Provides access to all clusters to users
- all transactionalid: Provides transactionalid access to users
- all delegationtoken: Provides delegationtoken access to users

You need to ensure that the required workload username or user group, and the ssb service user is added to the policies of the Kafka service and to the policies of the created Streams Messaging and Streaming Alaytics clusters.

	+ 🛛 🖸
cm_kafka	
csa_flink_cluster5_kafka_0d8c	
csm_kafka_cluster_1_kafka_d478	

Procedure

1. Click cm_kafka under Kafka service on the Service Manager page.

You are redirected to the cm_kafka Policies page.

Ŕ	Ranger	C Access Manager	🗅 Audit	Security Zone	Settings					ů	
	Service Manager) cm: kafka Policies Last Response Time : 02/02/2023 04:15:44 PM										
L	ist of Policies :	cm_kafka									
	Q. Search for your policy O Add New Police										
	Policy ID 🛦	Policy Name		Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action	
	25	all - consumergroup			Enabled	Enabled		_c_ranger_admins_21847a7a	cruisecontrol ssb streamsmsgmgr kafka + More	• 6	
	26	all - topic			Enabled	Enabled		_c_ranger_admins_21847a7a	cruisecontrol ssb streamsmsgmgr kafka + Nore		
	27	all - transactionalid		-	Enabled	Enabled		_c_ranger_admins_21847a7a	cruisecontrol ssb streamsmsgmgr kafka + More		
	28	all - cluster			Enabled	Enabled		_c_ranger_admins_21847a7a	cruisecontrol ssb streamsmsgmgr kafka + More		
	29	all - delegationtoken			Enabled	Enabled		_c_ranger_admins_21847a7a	cruisecontrol ssb streamsmsgmgr kafka + More	• 7	

- 2. Click on the edit button of the *all-consumergroup* policy.
- 3. Add the user group to the Select Group field under the Allow Conditions settings.
 - Alternatively, you can also add the workload usernames to the Select User field under the **Allow Conditions** setting.
- 4. Add the ssb service user to the Select User field under the Allow Conditions setting, if it is not configured by default.

🕅 Ranger	♥ Access Manager [🗅 Audit	f Security Zone 🛛 🌣	Settings				۵	
Service Manager	> cm_kafka Policies > Edit	Policy					Las	t Response Time : 02/02/20	23 05:01:56 PM
Allow Conditi	ons:								hide 🔶
	Select Role		Select Gr	roup	Select User	Policy Conditions	Permissions	Delegate Admin	
Select P	oles		×_c_ranger_admins_21847a	17a	× cruisecontrol [× ssb] × streamsmsgmgr × kafka [× kafka_mirror_maker] × streamsrepmgr	Add Conditions	Consume Describe Delete	Ø	×
Select R	oles		Select Groups		× rangerlookup	Add Conditions	Describe 🖋		×

5. Click Save.

You are redirected to the list of Kafka policies page.

- 6. Click on + More... to check if the needed workload user is listed under the Users for the edited policy. Repeate same steps to the remaining Kafka policies as well based on the required authorization level:
 - all topic
 - all transactionalid
 - all cluster

Configuring Schema Registry policies

After accessing the Ranger Admin Web UI, the workload username or user groups, and the SSB service user needs to be added to the Schema Registry policies to be able to use Flink and SSB.

About this task

The following resource based policy need to be configured for the Schema Registry service:

- all export-import: Provides import and export access for users.
- all serde: Provides access to store metadata for the format of how the data should be read and written.
- all schema-group, schema-metadata: Provides access to the schema groups, schema metadata, and schema branch.
- all schema-group, schema-metadata, schema-branch: Provides access to the schema groups, schema metadata, and schema branch.
- all schema-group, schema-metadata, schema-branch, schema-version: Provides access to schema groups, schema metadata, schema branch, and schema version.
- all registry-service: Provides access to the schema registry service, the user can access all Schema Registry entities.

You need to ensure that the required workload username or user group, and the ssb service user is added to the policies of the created Streams Messaging clusters.



Procedure

1. Select your Streams Messaging cluster under the **Schema Registry** service on the **Service Manager** page. You are redirected to the list of Schema Registry policies page.

rvice Manager	csm_kafka_cluster_1_schemaregistry Polic	ies	t connigo				Last Respon	se Time : 02/02/2023 05:15:09
of Policies : Q. Search for y	csm_kafka_cluster_1_schemaregist	ry					0) 0	Add New Police
Policy ID 🔺	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
116	all - export-import	-	Enabled	Enabled		_c_ranger_admins_21847a7a	ssb streamsmsgmgr kafka schemaregistry + More	• 7
117	all - serde	-	Enabled	Enabled		_c_ranger_admins_21847a7a	ssb streamsmsgmgr kafka schemaregistry + More	• 7
118	all - schema-group, schema-metadata	-	Enabled	Enabled		_c_ranger_admins_21847a7a	ssb streamsmsgmgr kafka schemaregistry + More	• 7
119	all - schema-group, schema-metadata, s	-	Enabled	Enabled		_c_ranger_admins_21847a7a vett-dfx-service-account-group	ssb streamsmsgmgr kafka schemaregistry + More	• 7 •
120	all - registry-service	-	Enabled	Enabled		_c_ranger_admins_21847a7a vett-dfx-service-account-group	ssb streamsmsgmgr kafka schemaregistry + More	• 7
121	all - schema-group, schema-metadata, s	-	Enabled	Enabled		_c_ranger_admins_21847a7a	ssb streamsmsgmgr kafka schemaregistry + More	• 7 1

- 2. Click on the edit button of the all-schema-group, schema-metadata, schema-branch, schema-version policy.
- 3. Add the user group to the Select Group field under the Allow Conditions settings.
 - Alternatively, you can also add the workload usernames to the Select User field under the **Allow Conditions** setting.
- **4.** Add the ssb service user to the Select User field under the **Allow Conditions** setting, if it is not configured by default.

Ranger	V Access Manager	Audit 🕤 Security Zone 🔹 Settings				💑 c	sso_dmars
Service Manager	csm_kafka_cluster_1_schemail	registry Policies 🔪 Edit Policy			Las	t Response Time : 02/02/20	023 05:22:38 F
Allow Conditio	ins:						hide 🔺
	Select Role	Select Group	Select User	Policy Conditions	Permissions	Delegate Admin	
Select Ro	les	x_c_ranger_admins_21847a7a	× ssb × streamsnsgmgr) × kafka × schemaregistry × rangerlookup	Add Conditions	Create Read Update Delete	2	×

5. Click Save.

You are redirected to the list of **Schema Registry policies** page.

6. Click on + More... to check if the needed workload user is listed under the Users for the edited policy.

Repeate same steps to the remaining Schema Registry policies as well based on the required authorization level:

- all export-import
- all serde
- all schema-group, schema-metadata
- all schema-group, schema-metadata, schema-branch, schema-version
- all registry-service

Configuring Hive policies

After accessing the Ranger Admin Web UI, the workload username or user groups, and the SSB service user needs to be added to the Hadoop SQL policies to be able to use Flink and SSB with Hive.

About this task

The following resource based policy need to be configured for the Hive service:

- all global: Provides global access to users.
- all database, table, column: Provides access to all databases, tables, and columns for users.
- all database, table: Provides access to all databases and tables for users.
- all database: Provides access to all databases for users.
- all hiveservice: Provides hiveservice access to users.
- all database, udf: Provides database and udf access to users.
- all url: Provides url access

You need to ensure that the required workload username or user group, and the ssb service user is added to the policies of the Hive service and the created Operational Database clusters.

	+ 🛛 🖸
Hadoop SQL	 Image: Contract of the second secon

Procedure

- 1. Click Hadoop SQL under Hadoop SQL service on the Service Manager page.
 - You are redirected to the list of **Hadoop SQL** policies page.

Ranger	🛡 Access Manager 📑 Audit	Security Zone	Settings					🙀 csso_dmarsa
Service Manager Access M ist of Policies :	Hadoop SQL Policies lasking Row Level Filter Hadoop SQL						Last Response	a Time : 02/02/2023 05:15:25 P
Q. Search for y	rour policy						• •	Add New Policy
Policy ID 🔺	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
8	ali - global	-	Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• • •
9	all - database, table, column		Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• •
10	all - database, table	-	Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• 7 0
11	all - storage-type, storage-url	-	Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• •
12	all - database	-	Enabled	Enabled		_c_ranger_admins_21847a7a public	hive beacon dpprofiler hue + More	• 7
13	all - hiveservice	-	Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• • •
14	ali - database, udf	-	Enabled	Enabled		_c_ranger_admins_21847a7a	hive beacon dpprofiler hue + More	• • •
15	all - uri		Enabled	Enabled	-	_c_ranger_admins_21847a7a	hive bescon dpprofiler hue + More	• 7

- 2. Click on the edit button of the *all-global* policy.
- 3. Add the user group to the Select Group field under the Allow Conditions settings.
 - Alternatively, you can also add the workload usernames to the Select User field under the **Allow Conditions** setting.
- 4. Add the ssb service user to the Select User field under the Allow Conditions setting, if it is not configured by default.

anger 🛛 Access Manager 🗋	Audit 🕑 Security Zone 🌣 Settings			ŵ
ce Manager > Hadoop SQL Policies > Edi	t Policy		Las	Response Time : 02/02/2023 05:19:
ow Conditions:				hide
Select Role	Select Group	Select User	Permissions	Delegate Admin
Select Roles	×_o_ranger_admins_21847a7a	x hive x beacon x dpprofiler x hue x admin x impala x ssb	select update Create Drop Atter Index Lock All Read Write ReplAdmin Service Admin Temporary UDF Admin Refresh RW Storage	×
Select Roles	Select Groups	× rangerlookup	Read solect	×

5. Click Save.

You are redirected to the list of Hadoop SQL policies page.

6. Click on + More... to check if the needed workload user and the ssb are listed under the Users for the edited policy.

Repeate same steps to the remaining Hadoop SQL policies as well based on the required authorization level:

- all-database, table, column
- all-database, table
- all-database
- all-database, udf
- all-hiveservice
- all-url

Configuring Kudu policies

After accessing the Ranger Admin Web UI, the workload username or user groups, and the SSB service user needs to be added to the Kudu policies to be able to use Flink and SSB.

About this task

The following resource based policy need to be configured for the Kudu service:

- all database, table: Provides access to all databases and tables for users.
- all database, table, column: Provides access to all databases, tables, and columns for users.
- all database: Provides access to all databases for users.

You need to ensure that the required workload username or user group, and the ssb service user is added to the policies of the Kudu service and the created Real-Time Data Mart clusters.



Procedure

1. Click cm_kudu under Kudu service on the Service Manager page.

You are redirected to the **cm_kudu Policies** page.

Ranger	🛡 Access Manager 🗋 Audit 🔮 S	ecurity Zone 🛛 🌣 Setti	ngs					🙀 csso_dmarsal
Service Manager	Cm_kudu Policies						Last Response	• Time : 02/02/2023 05:15:45 PM
st of Policies : o	cm_kudu							
Q Search for yo	ur policy					0		Add New Policy
Policy ID 🔺	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
86	all - database, table	-	Enabled	Enabled		_c_ranger_admins_21847a7a	kudu	• 7 •
87	all - database, table, column		Enabled	Enabled		_c_ranger_admins_21847a7a	kudu	• 7 •
88	all - database		Enabled	Enabled	-	_c_ranger_admins_21847a7a	kudu	• 7 •
insed under the Ap	ache License, Version 2.0							

- 2. Click on the edit button of the *all-database* policy.
- 3. Add the user group to the Select Group field under the Allow Conditions settings.
 - Alternatively, you can also add the workload usernames to the Select User field under the **Allow Conditions** setting.

4. Add the ssb service user to the Select User field under the Allow Conditions setting, if it is not configured by default.

langer 📭	Access Manager	🗋 Audit	🕈 Security Zone	Settings			ů	
rvice Manager > cr	m_kudu Policies 🔪	Edit Policy				Las	t Response Time : 02/02/2023	3 05:18:58
llow Conditions:								hide ^
	Select Role		Sele	ct Group	Select User	Permissions	Delegate Admin	
Select Roles	Select Role		Sele	ct Group 847a7a	Select User	Permissions SELECT INSERT UPDATE DELETE ALTER GREATE DROP METADATA ALL	Delegate Admin	×

5. Click Save.

You are redirected to the list of Kudu policies page.

- Click on + More... to check if the needed workload user is listed under the Users for the edited policy. Repeate same steps to the remaining Kudu policies as well based on the required authorization level:
 - all database, table
 - all database, table, column



Note: From CDP Public Cloud 7.2.16, Kudu environments have their own Ranger policy. Ensure that the permissions are added for every relevant Kudu cluster in the environment.