

Cloudera Stream Processing 2.0.0

Monitoring Kafka Cluster Replications using Streams Messaging Manager

Date published: 2019-09-13

Date modified: 2019-09-13

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with the letter 'E' in the middle of "UDERA" having a unique design where the top and bottom bars are separated by a horizontal gap.

<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

Monitoring Cluster Replications Overview.....	4
Viewing Replication Details.....	4
Searching Cluster Replications by Source.....	5
Monitoring Cluster Replications by Quick Ranges.....	5
Monitoring Status of the Clusters to be Replicated.....	6
Monitoring Topics to be Replicated.....	6
Searching by Topic Name.....	7
Monitoring Throughput for Cluster Replication.....	7
Monitoring Replication Latency for Cluster Replication.....	8
Monitoring Checkpoint Latency for Cluster Replication.....	9
Monitoring Throughput and Latency by Values.....	10

Monitoring Cluster Replications Overview

You can use Streams Replication Manager (SRM) to implement cross-cluster Kafka topic replication in SMM.

SRM is an enterprise-grade replication solution that enables fault tolerant, scalable and robust cross-cluster Kafka topic replication. SRM provides the ability to dynamically change configurations and keeps the topic properties in sync across clusters at high performance. SRM also delivers custom extensions that facilitate installation, management and monitoring making SRM a complete replication solution that is built for mission critical workloads. For more information about SRM, see the *Streams Replication Manager for HDF and HDP* document library.

You need to configure SRM properties in SMM. After you configure SRM in SMM, you can use SMM to monitor all the Kafka cluster replications available in your environment. You can view all replication details in SMM. You can monitor status of the Kafka cluster replications, number of topics associated with the replication, throughput, replication latency, and checkpoint latency of Kafka cluster replications in SMM.



Note: You can also create alerts to receive notifications based on the conditions that you configure in the alert policy for monitoring Kafka cluster replications in your system. For more information about creating alerts to monitor Kafka cluster replications, see the *Managing Alert Policies* guide.

Viewing Replication Details

After you select a cluster, SMM displays the details of the replications where the selected cluster is the destination cluster.

About this task

You can view the status of the replications, source cluster names, target cluster names, number of topics to be replicated, number of consumer groups, throughput of the replication, replication latency, and checkpoint latency. SMM also displays two graphs, one represents the throughput of the replication and the other displays the replication latency of the replication along with the details of each topic to be replicated.

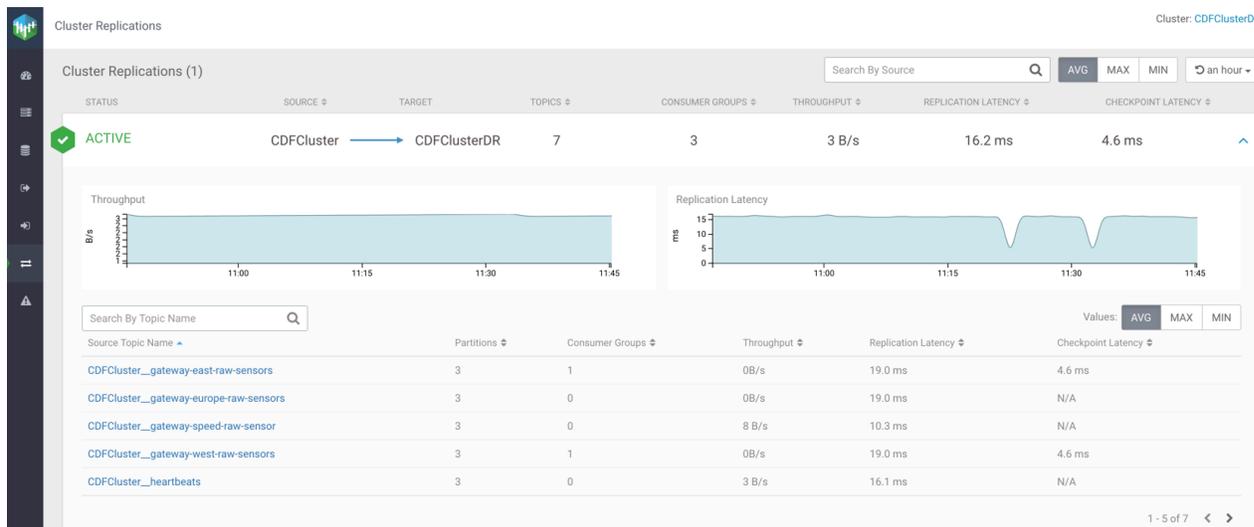
Perform the following steps to view the details of a cluster replication:

Procedure

In the Cluster Replications page, click a cluster replication or the drop-down icon beside a replication, as shown in the following image:

STATUS	SOURCE	TARGET	TOPICS	CONSUMER GROUPS	THROUGHPUT	REPLICATION LATENCY	CHECKPOINT LATENCY
ACTIVE	CDFCluster	CDFClusterDR	7	3	3 B/s	16.2 ms	4.6 ms

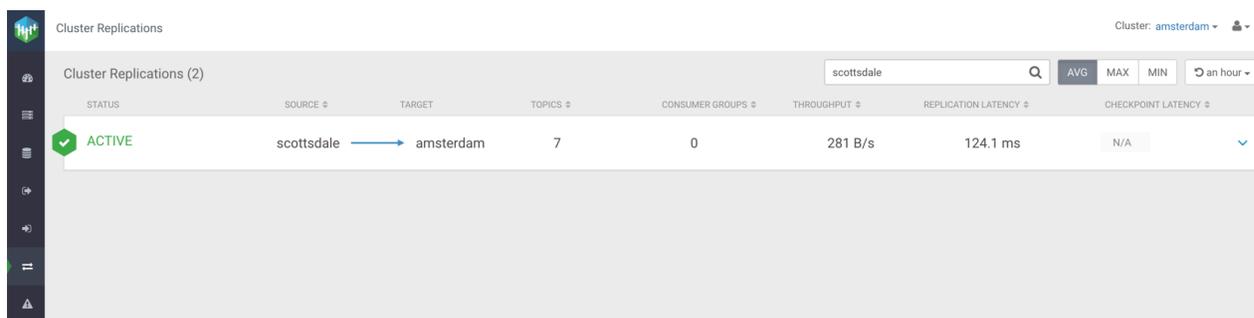
The replication details appear as shown in the following image:



Searching Cluster Replications by Source

About this task

You can use the Search By Source bar at the top right of the page to search cluster replications by the source cluster name. For example, if the destination cluster is amsterdam, and you enter scottsdale in the Search By Source bar, SMM displays the Kafka cluster replication between scottsdale and amsterdam, as shown in the following image:



Monitoring Cluster Replications by Quick Ranges

About this task

You can filter the cluster replications by time range. You can choose any of the following filter values from the drop-down to display the cluster replication details in the selected time range:

- Last 1 Hour
- Last 6 Hours
- Last 24 Hours
- Last 2 Days

The following image shows the Quick Ranges drop-down:

STATUS	SOURCE	TARGET	TOPICS	CONSUMER GROUPS	THROUGHPUT	REPLICATION LATENCY	CHECKPOINT
ACTIVE	springfield	amsterdam	9	0	745 B/s	8.0 ms	N/A
ACTIVE	scottsdale	amsterdam	7	0	281 B/s	125.3 ms	N/A

Monitoring Status of the Clusters to be Replicated

You can monitor the status of the Kafka cluster replications in the Status column in the Cluster Replications page.

The status has three variations:

- Active. Indicates that the cluster replication is running.
- Inactive. Indicates that the cluster replication is not running.
- Warning. Indicates that the cluster replication is facing issues.

If the status of a Kafka cluster replication shows inactive or warning, check the logs, and troubleshoot the replication.

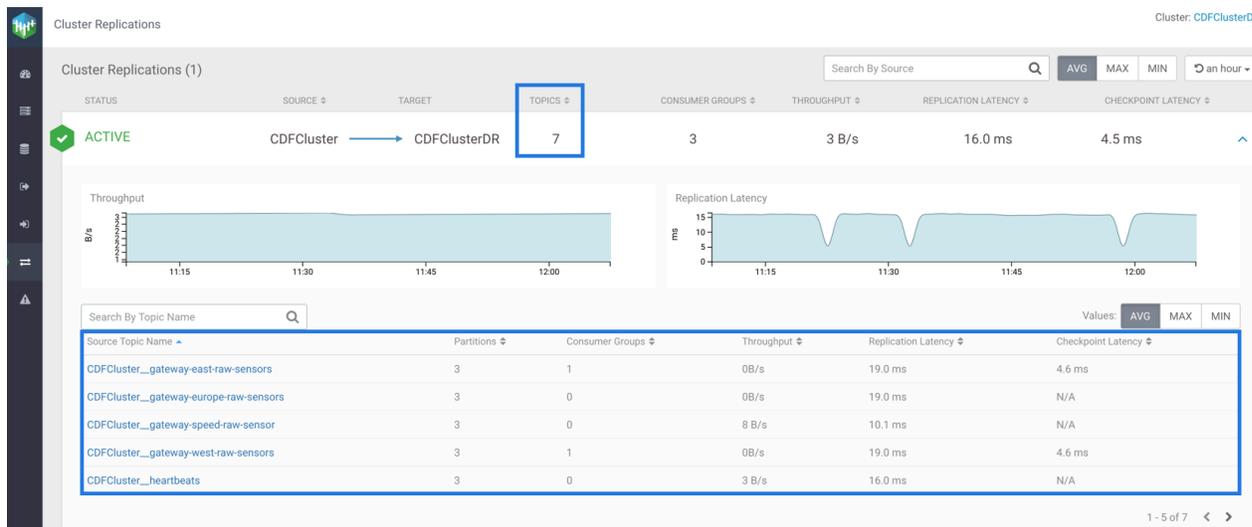
Monitoring Topics to be Replicated

You can monitor the number of topics associated with the replication in the Topics column in the Cluster Replications page.

Click on the cluster replication to fetch topic details. SMM displays the following details about the topics:

- Source Topic Name. Name of the topic at the source.
- Partitions. Number of partitions of the topic at the source.
- Consumer Groups. Number of consumer groups consuming data from the topic.
- Throughput. Data replicated between the source cluster and the target cluster per second from a topic. Throughput is measured in bytes per second. By default, SMM displays the average throughput. You can also fetch the maximum or minimum throughputs for a topic by clicking the MAX or MIN button located above the topic details.
- Replication Latency. Amount of time taken for a message of a topic to get replicated from the source cluster to the target cluster. Replication latency is measured in milliseconds. By default, SMM displays the average replication latency. You can also fetch the maximum or minimum replication latency for a topic by clicking the MAX or MIN button located above the topic details.
- Checkpoint Latency. Amount of time taken for a message of a topic to checkpoint on the target cluster after the message is committed on the source cluster. Checkpoint latency is measured in milliseconds. By default, SMM displays the average checkpoint latency. You can also fetch the maximum or minimum checkpoint latency for a topic by clicking the MAX or MIN button located above the topic details.

The following diagram shows details of topics in a cluster replication:



In the image, you can see that the number of topics to be replicated from CDFCluster to CDFClusterDR is 7, and topic details including topic names, number of partitions for the topics at the source cluster, number of consumer groups consuming messages from each topic, throughput, replication latency, and checkpoint latency of each topic.

Searching by Topic Name

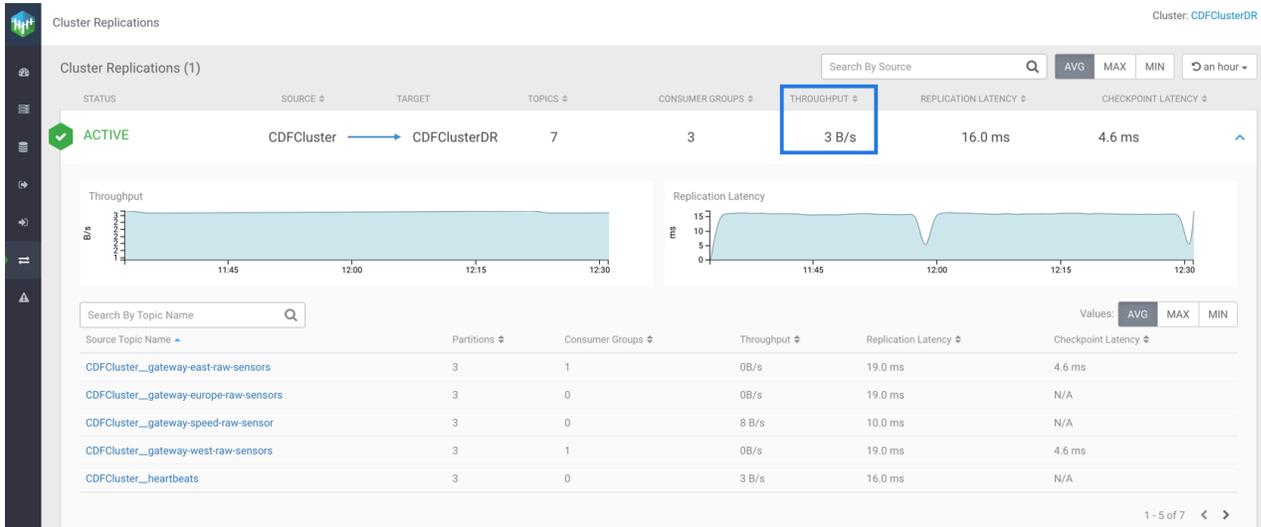
You can use the Search By Topic Name bar to search topic by names and get details of that topic. The following image shows details of the CDFCluster__heartbeats topic:



Monitoring Throughput for Cluster Replication

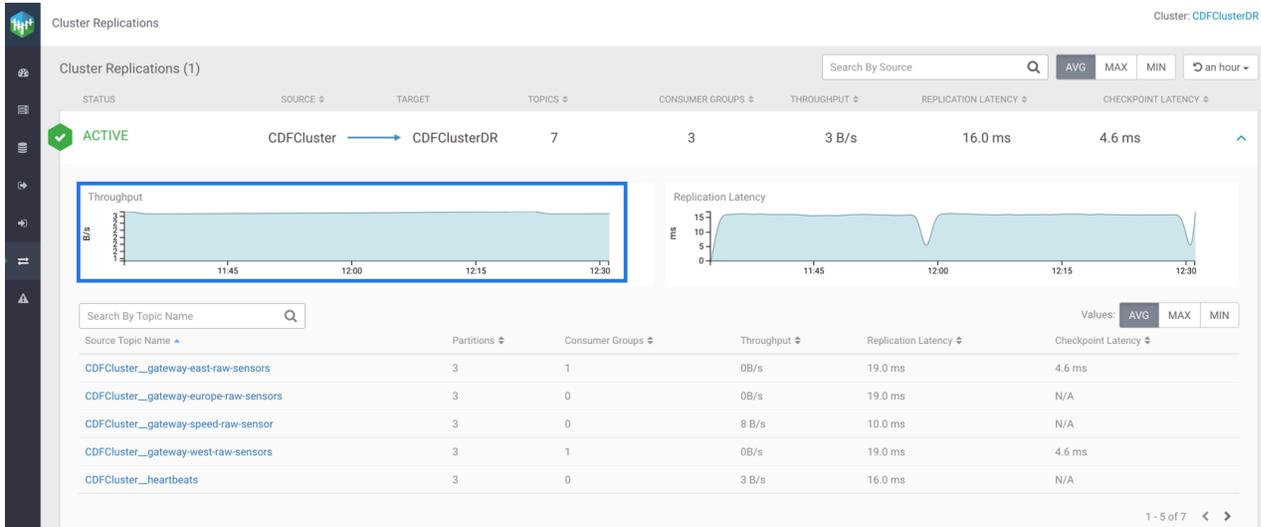
You can monitor the throughput of a Kafka cluster replication in SMM.

Throughput is defined as the data replicated between the source cluster and the destination cluster per second. Throughput is measured in bytes per second.



In the image, you can see the average throughput for the CDFCluster to CDFClusterDR replication is 3 bytes per second. You can fetch the maximum or minimum throughput for the replication by clicking the MAX or MIN button located above the cluster replication.

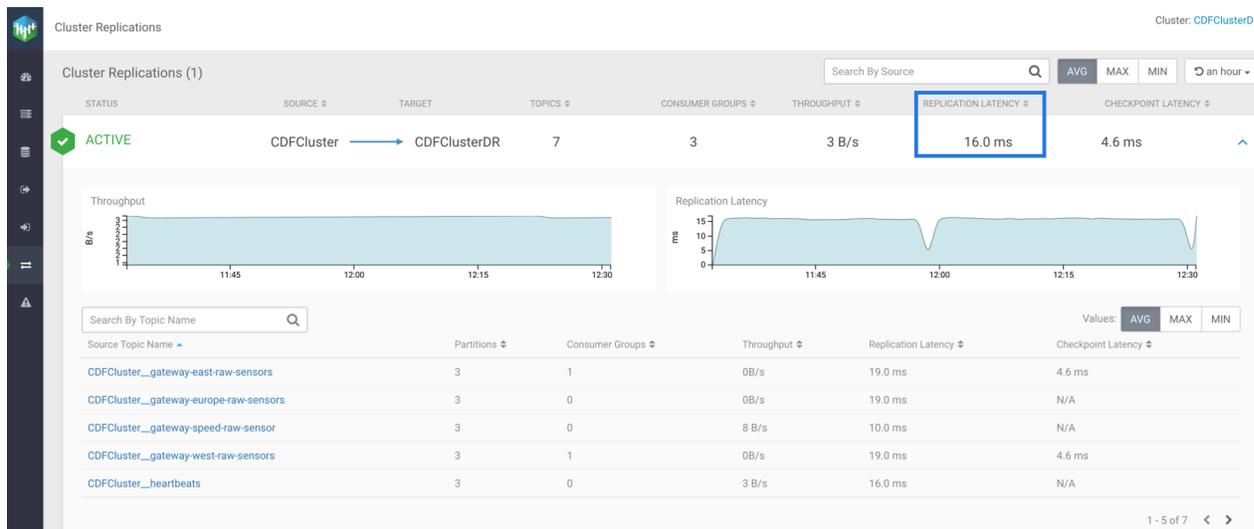
You can also monitor the throughput of a cluster replication graphically. SMM displays the Throughput graph for each cluster in the cluster details. The following image shows the graph for throughput for CDFCluster to CDFClusterDR replication:



Monitoring Replication Latency for Cluster Replication

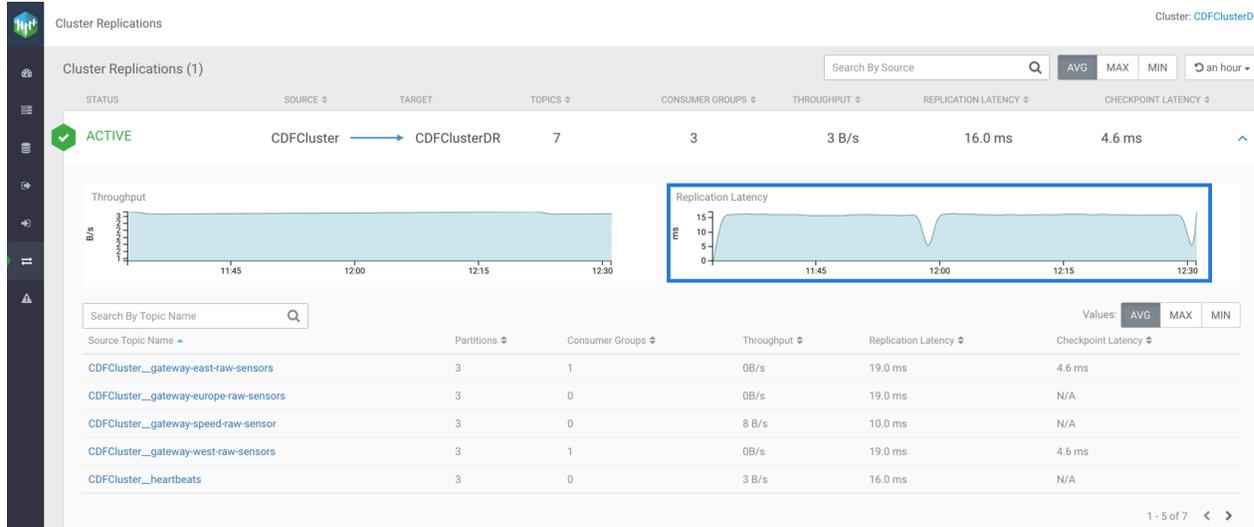
You can monitor the replication latency of a Kafka cluster replication in SMM.

Replication latency is defined as the amount of time taken for a message to get replicated from the source cluster to the destination cluster. Replication latency is measured in milliseconds.



In the image, you can see the average replication latency for the CDFCluster to CDFClusterDR replication is 16.0 milliseconds. You can fetch the maximum or minimum replication latency for the replication by clicking the MAX or MIN button located above the cluster replication.

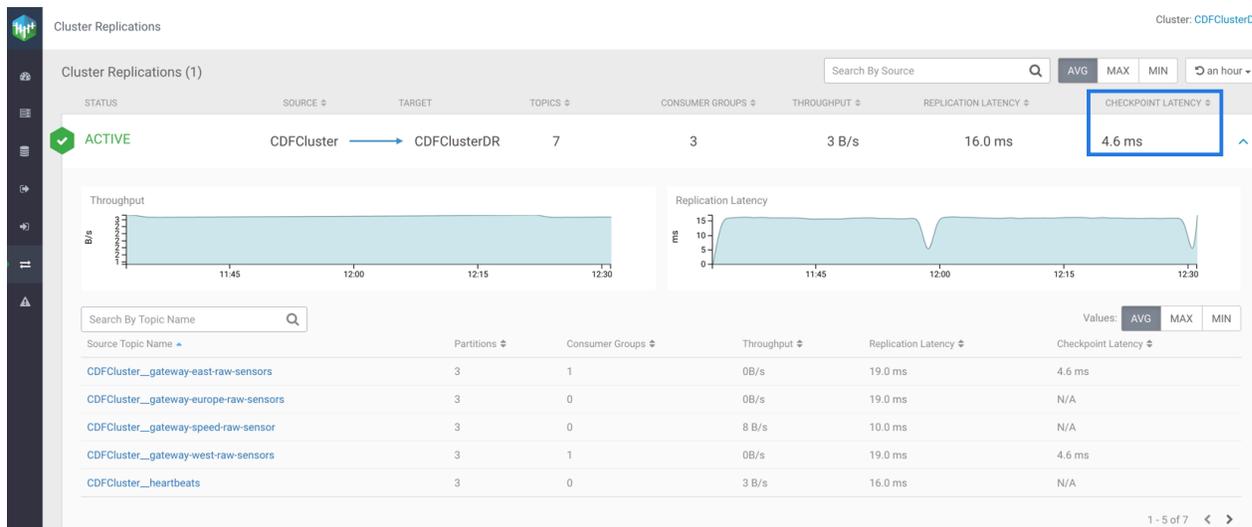
You can monitor the replication latency of a cluster replication graphically. SMM displays the Replication Latency graphs for each cluster in the cluster details. The following image shows the graph for replication latency for CDFCluster to CDFClusterDR replication:



Monitoring Checkpoint Latency for Cluster Replication

You can monitor the checkpoint latency of a Kafka cluster replication in SMM.

Checkpoint latency is defined as the amount of time taken for a message of a topic to checkpoint on the target cluster after the message is committed on the source cluster. Checkpoint latency is measured in milliseconds.



In the image, you can see the average checkpoint latency for the CDFCluster to CDFClusterDR replication is 4.6 milliseconds. You can fetch the maximum or minimum replication latency for the replication by clicking the MAX or MIN button located above the cluster replication.



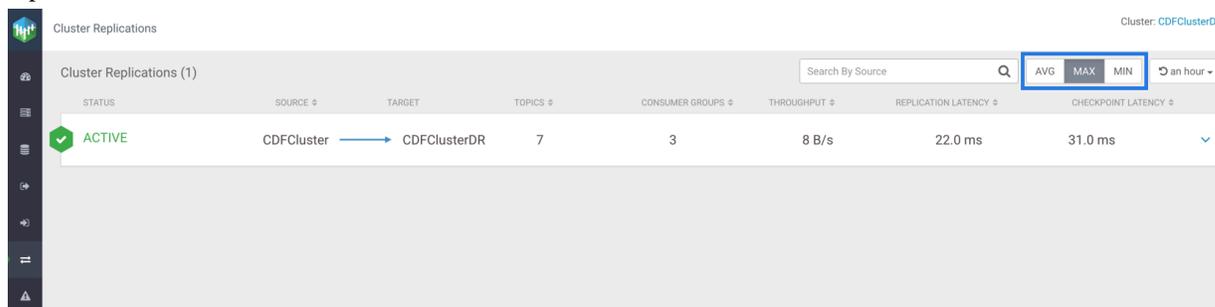
Note: If the checkpoint latency for a cluster replication or a topic appears as Not Available, then that means there are no consumer groups defined.

Monitoring Throughput and Latency by Values

You can fetch the average, maximum, and minimum values for throughput, replication latency, and checkpoint latency.

You can do this at the following levels:

- Cluster Replication Level. Click the AVG, MAX, or MIN buttons, as shown in the following image, to fetch average, maximum, or minimum values of throughput, replication latency, and checkpoint latency for cluster replications.



- Topic Level. Click the AVG, MAX, or MIN buttons, as shown in the following image, to fetch average, maximum, or minimum values of throughput, replication latency, and checkpoint latency for topics.

The screenshot displays the Cloudera Stream Processing interface for monitoring Cluster Replications. The main view shows a replication from CDFCluster to CDFClusterDR with 7 topics, 3 consumer groups, a throughput of 3 B/s, a replication latency of 16.0 ms, and a checkpoint latency of 4.6 ms. Two line graphs show throughput (B/s) and replication latency (ms) over time. Below the graphs is a table of replication details with columns for Source Topic Name, Partitions, Consumer Groups, Throughput, Replication Latency, and Checkpoint Latency. The 'Values' section at the top right of the table has buttons for AVG, MAX, and MIN, with the MIN button highlighted by a red box.

Source Topic Name	Partitions	Consumer Groups	Throughput	Replication Latency	Checkpoint Latency
CDFCluster__gateway-east-raw-sensors	3	1	0B/s	19.0 ms	2.0 ms
CDFCluster__gateway-europe-raw-sensors	3	0	0B/s	19.0 ms	N/A
CDFCluster__gateway-speed-raw-sensor	3	0	4 B/s	8.0 ms	N/A
CDFCluster__gateway-west-raw-sensors	3	1	0B/s	19.0 ms	2.0 ms
CDFCluster__heartbeats	3	0	0B/s	8.0 ms	N/A