

Cloudera Stream Processing Community Edition 2.0.0

Release Notes

Date published: 2022-07-29

Date modified: 2023-01-31

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

What's new in CSP Community Edition.....	4
January 2023.....	4
Streams Messaging.....	4
Streaming Analytics.....	6
July 2022.....	7
Component support.....	7
Limitations.....	8

What's new in CSP Community Edition

Learn about the new features of Cloudera Stream Processing (CSP) Community Edition.

January 2023

Notable changes, improvements, and new features for Streams Messaging and Streaming Analytics components in the January 2023 release.

Streams Messaging

What's new in Kafka, Kafka Connect, Schema Registry and Streams Messaging Manager.

Kafka and Kafka Connect

Kafka rebase to 3.1.2

Kafka shipped in this version of CSP Community Edition is based on Apache Kafka 3.1.2 (previous version was 2.8.1). For more information, see the official [Apache Kafka documentation](#).

New Kafka Connect connectors

The following new Kafka connect connectors are introduced:

- HDFS Stateless Sink
- Influx DB Sink
- Debezium Db2 Source



Important: If you want to deploy an instance of the Debezium Db2 Source connector, ensure that you first install the Db2 JDBC driver. For more information, see [Setting up connectors that use JDBC drivers](#).

Syslog TCP Source connector 2.0.0.

The Syslog TCP Source Kafka Connect connector is updated to version 2.0.0. The following notable changes and improvements are made:

- Three new properties are added, these are as follows:
 - Max Batch Size

This property controls the maximum number of messages to add to a single batch of messages. This is a required property. Its default value is 1.

- Authorized Issuer DN Pattern and Authorized Subject DN Pattern

These properties allow you to enable authorization for incoming TLS connections. Both properties accept regular expressions as a value. The configured regular expressions are applied against the Distinguished Names of incoming TLS connections. If the Distinguished Names do not match the pattern, the following message is logged and the messages do not get forwarded to Kafka.

```
Error: authorization failure
```

Both properties are optional and are set to .* by default.

- The Max Number of TCP Connections property is replaced by the Max Number of Worker Threads property.

Similarly to Max Number of TCP Connections, Max Number of Worker Threads is also used to specify the number of TCP connections, but instead of exactly specifying the number of allowed connections, you now specify how many worker threads are reserved for TCP

connections. Note that a single worker thread is capable of handling multiple connections. This is a required property. Its default value is 2.

- Existing version 1.0.0. connectors will continue to function. Upgrading them, however, is not possible. If you want to use the new version of the connector, you must deploy a new instance of the connector.
- Deploying a version 1.0.0. instance of the connector is no longer possible.

AvroConverter support for KConnect logical types

The AvroConverter now converts between Connect and Avro temporal and decimal types.

Expose log directory total and usable space through the Kafka API

[KAFKA-13958](#) is backported in Kafka shipped with this version of CSP Community Edition. As a result, the Kafka API now exposes metrics regarding the total and usable disk space of log directories. The information on log directory space is collected by SMM and is exposed on the SMM UI. Specifically, you can now view the current log size of topics as well as the total log size and remaining storage space of brokers. For more information on how you can monitor log size metrics on the SMM UI, see [Monitoring log size information](#) in the Cloudera Runtime library.

Schema Registry

Support for alternative jersey connectors in SchemaRegistryClient

`connector.provider.class` can be configured in Schema Registry Client. If it is configured, `schema.registry.client.retry.policy` should also be configured to be different than default.

This also fixes the issue with some 3rd party load balancers where the client is expected to follow redirects and authenticate while doing that.

Schema Registry CDC support - change default schema compatibility

When a new Avro schema is created and its compatibility is not explicitly set, then a default compatibility value is used. Until now, that value was always BACKWARD. After this change, users on the server side can configure the default value.

Streams Messaging Manager

Improved alter topic functionalities

You can now increase the number of partitions of a topic (but not decrease). The option is available on the Configs tab on the Topic Details page.

Partition Assignment tab on the Topic Details page

The Assignment tab, on the topic details page, shows the current state of the partitions and replicas of the topic. It shows some topic-level statistics and the replica assignment of all partitions.

Added sorter functionality to partition lists in SMM UI

The SMM UI contains Brokers and Topics pages where records contain broker or topic specific partition lists and their profile pages as well. All partition list columns become sortable.

Improved SMM UX for Kafka Connectors configuration

The connector selection and connector configuration workflow steps are now separated into two different steps. Search and autocomplete are now available for Connect configuration keys. The help icon provides detailed information about each configuration key. The data type of the configuration values can be chosen from the options menu. For more information, see [Setting connector configurations](#) in the Cloudera Runtime library.

Added partition log-size information to the SMM UI

The SMM UI shows log-size related information about brokers, topics, and partitions. Furthermore, warning messages appear when log directory related errors are reported by Kafka.

SMM connector profile shows connector level error

Errors causing the whole connector to fail are now displayed on the connector profile page if available.

Data explorer allows specifying consumer isolation.level

`"/api/v1/admin/topics/{topicName}/partition/{partitionId}/payloads"` endpoint has a new parameter: `"consumerIsolationLevel"`. The accepted values are `"read_committed"` and `"read_uncommitted"`. This sets the `"isolation.level"` config for the `KafkaConsumer` used for retrieving messages. The default value is `"read_uncommitted"`. The isolation level can also be configured using a newly introduced drop-down menu available on the Data Explorer page of a topic.

The Data Explorer page has a new design

The Data Explorer now uses the “from offset” and “record limit” parameters to select an offset window to query.

SMM UI Data Explorer shows null values explicitly

Data Explorer in the SMM UI now displays 'null' with italic style applied when the value is null rather than an empty value.

SMM uses a specific REST API to fetch list of topics

SMM Connect page now uses the Connect active topic tracking feature to list the topics used by the Connectors instead of the connector configuration's topic property. Sink Connectors show up based on which topics they consumed from (regardless of whether `"topics"` or `"topics.regex"` config was used), and Source Connectors show up based on which topics they produced into.

Improved configuration of SMM Kafka interceptors

New configuration prefix for SMM monitoring interceptor's background producer: `"smm.monitoring.interceptor.producer."`.

Clients that use either of the SMM monitoring interceptors (`MonitoringConsumerInterceptor`, `MonitoringProducerInterceptor`) use a background producer to push client metrics into Kafka every 30 seconds. This background producer now can be configured by passing Producer Configurations to the client that uses the interceptor with the `"smm.monitoring.interceptor.producer."` prefix. The prefix is trimmed and the remaining part of the configuration is passed to the background producer. For instance, if the user wishes to configure the `"batch.size"` property for the background producer, the following configuration should be passed: `"smm.monitoring.interceptor.producer.batch.size"`.

There is a different behavior, however, with the `"client.id"` property. If the user does not provide a configuration to the client id (`smm.monitoring.interceptor.producer.client.id`), the default is used, which is: `"smm-monitoring-interceptor"`.

SMM Data Explorer shows JSON output in pretty printed format

SMM Data Explorer can show JSON output in pretty printed format by using the "JSON Pretty Print" deserializer.

Related Information

[Cloudera Runtime 7.2.16 Release Notes](#)

Streaming Analytics

What's new in SQL Stream Builder.

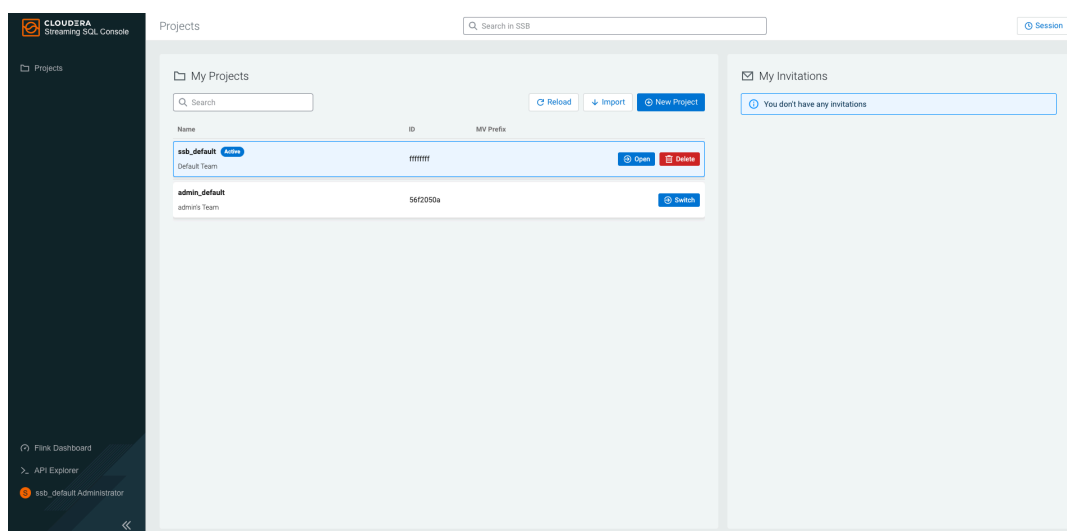
Apache Flink upgrade

Apache Flink 1.15.1 is supported in the Streaming Analytics 7.2.16 cluster definition.

For more information on what is included in the Apache Flink 1.15 version, see the [Apache Flink 1.15.1 Release Post](#) and the [Apache Flink 1.15 Release Notes](#)

Reworked Streaming SQL Console

The User Interface (UI) of SQL Stream Builder (SSB), the Streaming SQL Console has been reworked with new design elements.



Software Development Lifecycle (SDLC) support

Projects are introduced as an organizational element for SQL Stream Builder that allows you to create and collaborate on SQL jobs throughout the SDLC stages with source control.



Note: The Source Control features are in Technical Preview and not ready for production deployment. Cloudera encourages you to explore these features in non-production environments and provide feedback on your experiences through the *Cloudera Community Forums*.

Schema Registry JSON support

JSON is added as a supported data format for Schema Registry.

Materialized View pagination

You can set a limit and offset for the results of Materialized View queries to filter the query results more easily when accessing them through the Query API.

Related Information

[CSP CE Overview](#)

[CSP CE Installation Guide](#)

July 2022

What's new in CSP Community Edition July 2022.

This is the GA release of CSP Community Edition.

The Community Edition of CSP makes developing stream processors easy as it can be done right from your desktop or any other development node. Analysts, data scientists, and developers can now evaluate new features, develop SQL-based stream processors locally using SQL Stream Builder powered by Flink, and develop Kafka Consumers/Producers and Kafka Connect Connectors, all locally before moving to production.

Component support

Learn about the supported product and component version in Cloudera Stream Processing (CSP) Community Edition.

Product version	Component version
Cloudera Runtime 7.2.16	Apache ZooKeeper 3.5.5
Streams Messaging 7.2.16	Apache Kafka 3.1.2

Product version	Component version
	Streams Messaging Manager 2.3.0
	Schema Registry 0.10.0
Streaming Analytics 7.2.16	Apache Flink 1.15.1

Related Information

[Cloudera Runtime documentation](#)

[Streams Messaging documentation](#)

[Streaming Analytics documentation](#)

Limitations

Learn about the limitations, and the impacts or changes to the functionality of components in Cloudera Stream Processing (CSP) Community Edition.



Important: CSP Community Edition is an unsecured deployment of Cloudera Streams Messaging and Streaming Analytics. The security features available in Apache Kafka, Apache Flink, SQL Stream Builder, Schema Registry, and Streams Messaging Manager are not supported. In addition, support services are not provided for the Community Edition. As a result, CSP Community Edition is not meant for the requirements to be used in production.

Streams Messaging

Kafka

Kafka producers have a quota of 100000 bytes per second. This means that producers get throttled if they reach this threshold.

Schema Registry

Schema Registry is limited to a maximum of 50 schemas.

Streaming Analytics

SQL Stream Builder

- Only 3 SQL jobs can be run simultaneously.
- Hive catalog is not supported.
- Iceberg integration is not supported.
- Logging customization under Advanced Settings is not supported.
- Job Notification is not supported.
- All jobs are submitted to the same standalone cluster.
- Storage option is only the docker volumes.