

Cloudera Streaming Analytics 1.10.0

Release Notes

Date published: 2019-12-17

Date modified: 2023-06-12

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 Cloudera Streaming Analytics.....	4
Fixed issues.....	4
Known issues and limitations.....	5
Behavioral changes.....	7
Unsupported features.....	8
Support Matrix.....	8
Component support.....	8
System Requirements.....	9
Default ports for Flink and SSB.....	10
Maven dependencies in Flink.....	10
Flink API Support.....	12

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.10.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

Rebase to Apache Flink 1.16

Apache Flink 1.16 is supported in Cloudera Streaming Analytics 1.10.0.

For more information on what is included in the Apache Flink 1.16 version, see the [Apache Flink 1.16 Release Post](#) and the [Apache Flink 1.16 Release Notes](#)

Built-in Data Visualization tool

The built-in data visualization tool in SQL Stream Builder (SSB) enables you to present the sampling data and the results of the Materialized View query using widgets on Streaming SQL Console. As the widgets are integrated into SSB, the visualization tool works out of the box without any dependencies, which offers easy access to the underlying, running jobs as data sources.

For more information, see the [Using widgets for data visualization](#) documentation.

Schema Handling

Cloudera Schema Registry catalog can now handle key schemas as well based on the provided key (and value) schema name suffix for custom naming convention.

For more information, see the [Adding Catalogs](#) documentation.

PyFlink Support

Python API for Flink is supported in Cloudera Streaming Analytics that enables you to create Flink DataStream applications using Python.

For more information, see the [Application development](#) and [Job lifecycle](#) documentation.

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.10.0.

CSA-3754 - The display name of the loadbalancer.url property should be "Load Balancer Host"

The issue regarding the loadbalancer.url property being duplicated in Cloudera Manager for SQL Stream Builder configuration is fixed.

CSA-4030 - Webhook sending fails when webhook template is empty string

The issue about creating a webhook table when the template is empty is fixed.

CSA-4333 - Use Kafka Timestamps switch reflects invalid value

The issue about incorrect status of the Kafka Timestamp configuration when viewing DDL is fixed.

CSA-4370 - Virtual tables imported from a schema in a Schema Catalog fail to describe correctly

The issue regarding the describe error for Virtual Tables using schemas in a Schema Registry Catalog is fixed.

CSA-4400 - Cannot delete invalid catalog

The issue regarding invalid catalogs can be registered, but cannot be deleted is fixed.

CSA-4412 - Cannot delete mv endpoint when it contains dynamic parameters

The issue regarding Materialized View endpoint cannot be deleted when using dynamic parameters is fixed.

CSA-4425 - Password in Kafka Data Source can be revealed after save

The issue regarding viewing the password after saving the Kafka Data Source is fixed.

CSA-4426 - Kafka Data Source name accepts spaces

The issue about accepting Kafka Data Sources with spaces in their name is fixed.

CSA-4427 - State of Execute and Stop options in Job context menu do not correspond the Job state

The issue regarding the invalid property in the upsert Kafka table template is fixed.

CSA-4428 - upsert-kafka template has an invalid property

The issue regarding the invalid property in the upsert Kafka table template is fixed.

CSA-4548 - Files cannot be uploaded through Swagger

The issue about the error when uploading files using REST API with Swagger is fixed.

CSA-4620 - Encryption for environment properties

The issue regarding encrypted environment properties is fixed.

CSA-4643 - flink-yarn-session is ignoring command line parameters

The issue regarding the ignored parameters that added to the flink-yarn-session in command line is fixed.

FLINK-18027 - ROW value constructor cannot deal with complex expressions

The fix regarding the Calcite error in FLINK-18027 is backported.

Known issues and limitations

Learn about the known issues in Flink and SQL Stream Builder, the impact or changes to the functionality, and the workaround in Cloudera Streaming Analytics 1.10.0.

SQL Stream Builder

CSA-4858 - Kerberos encryption type detection does not always work correctly for SSB

SSB detects no supported encryption types even though there is a list of allowed encryption types in the krb5.conf file. This causes an error when generating keytabs from the principal and password pair.

1. Run ktutil on your cluster.
2. Change the configuration with the following commands:

```
addent -password -p <username> -k 1 -e aes256-cts  
wkt /tmp/new_keytab.keytab
```

3. Upload the new keytab on Streaming SQL Console.

CSA-4800 - ToString of Job can cause stack overflow

The jobLogItems can cause stack overflow errors when toString is called.

None

CSA-4799 - Table Metadata is not saved when job is run via sql/execute

The SqlExecutorService.persistIfNeededAndExecute does not save the table metadata before executing the SQL job, therefore the data cleaner of the Materialized View Engine does not clean up the data based on the retention settings.

None

CSA-4699 - Keytab upload starts failing in SSB after some time, requiring a restart

The /tmp/ssb_keytab_work_dir is removed after a period of time and SSB can no longer create keytabs as the directory does not exist anymore.

Restart the SSB service to recreate the directory for the keytabs.

CSA-4650: Inconsistent sidebar collapse behavior

The sidebar is collapsed inconsistently on the homepage of Streaming SQL Console when opening a project.

None

Limitations when configuring widgets

The following widget configuration options are not available for certain widgets on Streaming SQL Console:

- Gauge visualization type: Expand on hover, Unit
- Donut visualization type: Expand on hover, Title
- Pie visualization type: Expand on hover

None

Upgrading to CSA 1.10 requires additional step

When upgrading to Streaming Analytics 1.10, the Update Admin Database step will fail during the upgrade process.

Select Repair Admin Database when the Update Admin Database process fails, and then complete the Update Admin Database process again.

Auto discovery is not supported for Apache Knox

You need to manually configure Knox with SQL Stream Builder to enable Knox authentication.

Complete the configuration based on the CDP Private Cloud Base version you use. For more information, see the [Enabling Knox authentication for SSB](#) documentation.

CSA-5006 - SSB service fails when using Active Directory (AD) Kerberos authentication

If you use AD Kerberos for authentication and the Load Balancer URL is not provided, it can cause the SQL Stream Builder (SSB) service to fail. The issue is caused by the keytab generation. When the keytab is generated by Cloudera Manager it requires the principals from the AD for the Load Balancer host, and without no host specified for the Load Balancer, the SSB service cannot be started by Cloudera Manager. This issue also persists when the Load Balancer role is not deployed or used with SSB.

Fill out the Load Balancer URL parameter in Cloudera Manager regardless of using Load Balancer with SSB. For more information, see the [Enabling High Availability for SSB](#) documentation.

Flink

In Cloudera Streaming Analytics, the following SQL API features are in preview:

- Match recognize
- Top-N
- Stream-Table join (without rowtime input)

DataStream conversion limitations

- Converting between Tables and POJO DataStreams is currently not supported in CSA.
- Object arrays are not supported for Tuple conversion.
- The java.time class conversions for Tuple DataStreams are only supported by using explicit TypeInformation: LegacyInstantTypeInfo, LocalTimeTypeInfo.getInfoFor(LocalDate/LocalDateTime/LocalTime.class).
- Only java.sql.Timestamp is supported for rowtime conversion, java.time.LocalDateTime is not supported.

Kudu catalog limitations

- **CREATE TABLE**
 - Primary keys can only be set by the `kudu.primary-key-columns` property. Using the `PRIMARY KEY` constraint is not yet possible.
 - Range partitioning is not supported.
- When getting a table through the catalog, `NOT NULL` and `PRIMARY KEY` constraints are ignored. All columns are described as being nullable, and not being primary keys.
- Kudu tables cannot be altered through the catalog other than simply renaming them.

Schema Registry catalog limitations

- Currently, the Schema Registry catalog / format only supports reading messages with the latest enabled schema for any given Kafka topic at the time when the SQL query was compiled.
- No time-column and watermark support for Registry tables.
- No `CREATE TABLE` support. Schemas have to be registered directly in the SchemaRegistry to be accessible through the catalog.
- The catalog is read-only. It does not support table deletions or modifications.
- By default, it is assumed that Kafka message values contain the schema id as a prefix, because this is the default behaviour for the SchemaRegistry Kafka producer format. To consume messages with schema written in the header, the following property must be set for the Registry client: `store.schema.version.id.in.header: true`.

Behavioral changes

Learn about the change in certain functionality of Flink and SQL Stream Builder (SSB) that has resulted in a change in behavior from the previously released version of Cloudera Streaming Analytics.

SQL Stream Builder

Summary:

Dynamic MV parameters have different configuration steps

Previous behavior:

Dynamic parameters were specified after the Materialized View query is submitted on the created endpoint URL.

New behavior:

Dynamic and static parameters can be selected when configuring the Materialized View query.

Summary:

Stateless SSB API

Previous behavior:

Requests from the SSB API could only be used for the active project that was selected.

New behavior:

The new v2 SSB API is based on the project ID that is specified in the endpoint.

Summary:

Restructuring Organization to Projects

Previous behavior:

The Organization ID was used and could be customized to as an identification method for SQL jobs and to authenticate users and teams.

New behavior:

The Organization ID was changed to Project ID due to the restructuring of SSB and Streaming SQL Console. For more information, see the [Projects](#) documentation.

Flink

Summary:

Additional step required when upgrading to latest version of CSA

Previous behavior:

Creating Job Result Store was not needed when upgrading to latest version of CSA.

New behavior:

When upgrading to latest version of CSA, you need to create the Job Result Store for Flink.

Unsupported features

The following features are not supported in Cloudera Streaming Analytics 1.10.0.

SQL Stream Builder

- Direct SQL Stream Builder upgrade from 1.3.0



Important: This does not impact Flink, you can directly upgrade Flink as described in the documentation.

For more information, see the [Upgrading SQL Stream Builder](#) in the 1.3.0 documentation.

Flink

- Apache Flink batch (DataSet) API
- GPU Resource Plugin
- Application Mode deployment
- SQL Client
- The following features are not supported in SQL and Table API:
 - HBase Table Connector
 - Old Planner
 - Non-windowed (unbounded) joins, distinct

Support Matrix

Before installing Cloudera Streaming Analytics, review the supported components, databases, connectors and the default ports in use for Flink and SQL Stream Builder (SSB).

Component support

Learn more about which Apache Flink component version is supported in the Cloudera Streaming Analytics (CSA) releases.

CSA version	Component version
CSA 1.10	Apache Flink 1.16.1
CSA 1.9.0	Apache Flink 1.15.1

CSA version	Component version
CSA 1.8.0	
CSA 1.7.0	
CSA 1.6.2	Apache Flink 1.14
CSA 1.6.1	
CSA 1.6.0	
CSA 1.5.3	Apache Flink 1.13
CSA 1.5.1	
CSA 1.5.0	
CSA 1.4.1	Apache Flink 1.12
CSA 1.4.0	
CSA 1.3.0	
CSA 1.2.0	Apache Flink 1.10
CSA 1.1.0	Apache Flink 1.9.1

Related Information

[CSA 1.9.0 Release Notes](#)

[CSA 1.8.0 Release Notes](#)

[CSA 1.7.0 Release Notes](#)

[CSA 1.6.2 Release Notes](#)

[CSA 1.6.1 Release Notes](#)

[CSA 1.6.0 Release Notes](#)

[CSA 1.5.3 Release Notes](#)

[CSA 1.5.1 Release Notes](#)

[CSA 1.5.0 Release Notes](#)

[CSA 1.4.1 Release Notes](#)

[CSA 1.4.0 Release Notes](#)

[CSA 1.3.0 Release Notes](#)

[CSA 1.2.0 Release Notes](#)

[CSA 1.1.0 Release Notes](#)

System Requirements

Before installing Cloudera Streaming Analytics, you should verify that you meet the system requirements. Other than CDP Private Cloud Base, you should also check the latest supported version of the needed components.

For detailed information about the supported versions of CDP Private Cloud Base, operating systems and databases, see the [Cloudera Support Matrix](#).

Apache Flink support	1.16.1	
Cloudera Runtime component support		
	7.1.7	7.1.8
Kafka ¹	2.5.0	3.1.1
Schema Registry	0.10.0	0.10.0

¹ Connecting to Kafka that is running on remote CDH6 or HDP3 is also supported.

Apache Flink support		1.16.1
Cloudera Runtime component support		
Streams Messaging Manager	2.1.0	2.3.0
HBase	2.2.3	2.4.6
HDFS	3.1.1	3.1.1
Atlas	2.1.0	2.2.0
Kudu	1.15.0	1.15.0
Hive	3.1.3	3.1.3

Connector support	
JDBC PostgreSQL	9.6-12
JDBC MySQL	5.7, 8
JDBC Hive	3.1.3
CDC PostgreSQL	9.6-12
CDC MySQL	5.7, 8
CDC Oracle	19.0.0
CDC Db2	11.5
CDC SQL Server	2007-2022

Default ports for Flink and SSB

You need to use the default ports of Flink and SSB when you need to reach or connect to their services. The default ports are set in Cloudera Manager, but can be changed if required.

The following table lists the default ports and the corresponding property file names for Flink and SQL Stream Builder (SSB). The ports are set by default in Cloudera Manager. You can change the ports as required using the configuration properties.

Component	Service	Port	Configuration property
Flink	Flink Dashboard	18211	historyserver.web.port
SQL Stream Builder	Streaming SQL Engine	18121	server.port
	Materialized View Engine	18131	server.port
SQL Stream Builder with Load Balancer	Streaming SQL Engine	8080	ssb.sse.loadbalancer.server.port
	Secured Streaming SQL Engine	8445	ssb.sse.loadbalancer.server.secure.port
	Materialized View Engine	8081	ssb.mve.loadbalancer.server.port
	Secured Materialized View Engine	8444	ssb.mve.loadbalancer.server.secure.port

For the default port list of the Cloudera Runtime components, see the *Ports Used by Cloudera Runtime Components* document.

Maven dependencies in Flink

Review the list of Maven dependencies to ensure the correct connector versions in your Flink applications.

Avro

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-avro</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

Confluent Registry

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-confluent-registry</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>

<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-avro-confluent-registry</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

CSV

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-csv</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

Hive

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-hive_2.12</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

HBase**Iceberg****JSON**

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-json</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

Kafka

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-kafka</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

Kudu

```
<dependency>
```

```
<groupId>org.apache.bahir</groupId>
<artifactId>flink-connector-kudu_2.12</artifactId>
<version>1.1.0-csa1.10.0.0</version>
</dependency>
```

Schema Registry

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-cloudera-registry</artifactId>
<version>1.16.1-csa1.10.0.0</version>
</dependency>
```

Table API

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-api-java-bridge</artifactId>
  <version>1.16.1-csa1.10.0.0</version>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-planner_2.12</artifactId>
  <version>1.16.1-csa1.10.0.0</version>
</dependency>
```

For more information about how to use Maven in Flink, see the [Apache documentation](#).

Flink API Support

Cloudera Streaming Analytics (CSA) offers support for three fundamental layers of the Apache Flink API. You can use DataStream API, the ProcessFunction API and a selected subset of the SQL API to develop your Flink streaming applications.

From the DataStream and ProcessFunction APIs, the following are supported based on the support annotations provided by the Apache Flink community.

Stable (@Public)	Evolving (@PublicEvolving)
<ul style="list-style-type: none"> DataStream API 	<ul style="list-style-type: none"> ProcessFunction Stream Join Interval Join Stateful operators FsStatebackend with HDFS RocksDBStateBackend with HDFS



Note: CSA does not support batch processing (DataSet API).