

Cloudera Streaming Analytics 1.11.0

## Release Notes

Date published: 2019-12-17

Date modified: 2023-09-12

# 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

<b>What's new in Cloudera Streaming Analytics.....</b>	<b>4</b>
<b>Fixed issues.....</b>	<b>4</b>
<b>Known issues and limitations.....</b>	<b>5</b>
<b>Behavioral changes.....</b>	<b>8</b>
<b>Unsupported features.....</b>	<b>9</b>
<b>Support Matrix.....</b>	<b>9</b>
Component support.....	10
System Requirements.....	10
Default ports for Flink and SSB.....	11
Maven dependencies in Flink.....	12
Flink API Support.....	13

## What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.11.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

### CSA 1.11.0.1

#### Hotfix for CSA 1.11.0

On September 22, 2023, CSA 1.11.0.1 was released as a hotfix to resolve some issues. If you have previously upgraded to CSA 1.11.0.0, upgrade to CSA 1.11.0.1 to access the latest version of CSA. You can find the new download links in the [Download location of CSA](#).

For more information about the issues fixed in CSA 1.11.0.1, see the [Fixed issues](#) section.

### CSA 1.11.0.0

#### Rebase to Apache Flink 1.16.2

Apache Flink 1.16.2 is supported in Cloudera Streaming Analytics 1.11.0.

For more information on what is included in the Apache Flink 1.16.2 version, see the [Apache Flink 1.16 Release Announcement](#).

#### Apache Iceberg support

Support for Apache Iceberg table has been added as a connector to Flink. This allows you to create Iceberg tables using the Streaming SQL Console.

For more information, see the [Creating Iceberg tables](#) and [Iceberg with Flink](#) documentation.

#### Support on CDP Private Cloud Base 7.1.9

CSA 1.11.0 is supported on CDP Private Cloud Base 7.1.9.

For more information about the supported operating system and database versions, see the [Support Matrix](#) page.

## Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.11.0.

### 1.11.0.1

#### CSA-4938 - Activating an environment makes SSB unable to start

The issue regarding the `NullPointerException` caused by the environment files 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.

#### CSA-4861 - Error with Flink JSON row serializer init in SSB

The issue regarding the `NullPointerException` NPE in SSB that is caused by not transferring the transient fields of `RowRowConverter` to the Flink workers is fixed.

### 1.11.0.0

#### CSA-4558 - SSB internal topics are not explicitly created

The issue about SQL Stream Builder (SSB) not creating Kafka topics for sampling is fixed, the topics are explicitly created even when the automatic topic creation is disabled.

#### CSA-4624 - Sensitive field redaction can be circumvented

The issue about accessing hidden or redacted fields of tables is fixed.

**CSA-4641 - Mask sensitive information in Environment**

The issue about accessing hidden or redacted fields of tables is fixed.

**CSA-4650 - Inconsistent sidebar collapse behavior**

The issue regarding the behavior of inconsistent sidebar collapse is fixed.

**CSA-4651 - Wrong padding for arrows on the right at the project homepage**

The issue regarding the incorrect padding for arrows on the **Projects** homepage is fixed.

**CSA-4670 - Default operator of MV query ruleset is not set on UI**

The issue about showing default operators not set in the Materialized View query ruleset is fixed.

**CSA-4672 - Fix chart wizard config options**

The issues regarding the Gauge chart are fixed. The units can be displayed, the Gauge chart width can be changed and colors also can be switched.

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

The issue regarding keytab upload failure in SSB after a period of time is fixed.

**CSA-4723 - Adding delete source setting button**

The issue about source settings cannot be deleted from Streaming SQL Console is fixed.

**CSA-4761 - UDF is created with an empty parameter type if none is specified**

The issue about creating UDFs with an empty parameter type that causes failure in job execution is fixed.

**CSA-4798 - Improve error handling of MV queries**

The issue about the dynamic parameter throwing an exception when not providing a value is fixed.

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

The issue about not saving table metadata when jobs are executed using REST API is fixed.

**CSA-4800 - ToString of Job can cause stack overflow**

The issue regarding stack overflow error caused by the jobLogItems field when calling toString is fixed.

**CSA-4806 - DataCleaner can miss Materialized View data to be cleaned**

The issue regarding Materialized View data not getting cleaned is fixed.

**CSA-4820 - Prevent projects with running jobs from being deleted**

The issue about running jobs is deleted when removing projects is fixed.

## 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 1.11.0.

### SQL Stream Builder

**CSA-5138 - SQL job submissions with UDF JARs fail when checkpointing is enabled**

Due to the handling of ClassLoaders for custom JARs, uploading any Java UDFs with checkpoints enabled will cause the SQL job to fail with the following error:

```
ERROR com.cloudera.ssb.sqlio.service.SqlExecutorService: Error while submitting streaming job
```

```
org.apache.flink.util.FlinkRuntimeException: org.apache.flink.
api.common.InvalidProgramException: Table program cannot be comp
iled.
```

Once the SQL job fails, the session on Streaming SQL Console must be reset before resubmitting the job without checkpointing.

None

### CSA-4960 - Invalid job schemas for existing SSB jobs

After successfully upgrading to CSA 1.11.0 from CSA 1.8.0 or lower versions, migrating the existing jobs produce invalid job schemas in the admin database.

The mv\_config object is stored in the mv\_config column of the SSB jobs table. You need to manually update the jobs table to resolve the issue:

- If an existing mv\_config includes an unknown create field, the field must be deleted.
- If an existing mv\_config includes the minRowRetentionCount deprecated key, the key should be changed to min\_row\_retention\_count.

The following mv\_config objects show an invalid and valid example:

- Valid mv\_config object:

```
{ "name": "quizzical_benz", "retention": 300, "min_row_reten
tion_count": 0, "recreate": false, "key_column_name": "", "a
pi_key": null, "ignore_nulls": false, "require_restart": fal
se, "enabled": false }
```

- Invalid mv\_config object:

```
{ "create": false, "name": "quizzical_benz", "retention": 300,
"minRowRetentionCount": 0, "recreate": false, "key_column_nam
e": "", "api_key": null, "ignore_nulls": false, "require_res
tart": false, "enabled": false }
```

### ENGESC-23078 - Job not found after successful job creation

After successfully creating a job in SSB, the SQL job is not found due to tables having empty values. This issue is indicated with the following error message in the log files:

```
java.lang.IllegalArgumentException: argument "content" is null
```

The issue only applies when upgrading from a CSA version lower than 1.9.0.

Update the empty values with null string in the mv\_config and checkpoint\_config fields as shown in the following example:

```
UPDATE jobs SET mv_config = 'null' WHERE mv_config IS NULL;
UPDATE jobs SET checkpoint_config = 'null' WHERE checkpoint_conf
ig IS NULL;
```

### CSA-4938 - Activating an environment makes SSB unable to start

Restarting SSB with an active environment file causes a NullPointerException and SSB fails to start.

Use the following command to update the environment file:

```
UPDATE environments SET properties = '{} ' WHERE properties IS NU
LL;
```



**Note:** This command results in removing the properties of the environment file.

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

When adding parameters to the Flink session using `flink-yarn-session -d` in command line, the parameters are not applied to the session.

None

#### 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-4861 - Error with Flink JSON row serializer init in SSB

As the `open()` function of the Flink `RowRowConverter` is called in SSB, the transient fields of the `RowRowConverter` are not transferred to the Flink workers that do the row serialization. This causes a `NullPointerException` (NPE). The error only occurs in case of composite fields (for example, `ARRAY`).

None

#### 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

#### FLINK-20539 - Type mismatch when using ROW() in computed column

Using the `ROW()` function in SQL statements fails due to a mismatch between Calcite and Flink SQL data type.

None

### Limitations

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:**

Kafka sink JSON serializer is refactored to the Flink implementation

Previous behavior:

An SSB specific implementation was used for the JSON serializer.

New behavior:

The `RowJsonSerializer` is added to be used for the Kafka sink JSON serializer. The v2 API version is also added for the `JSONSampler`.

#### **Summary:**

Remove user name length limitation

Previous behavior:

Username limitation was set to minimum four characters.

New behavior:



Limitation is removed for username length.

**Summary:**

Make API v2 the default selection in the SSB Swagger UI

Previous behavior:

Version could be selected for the REST API when using Swagger UI.

New behavior:

V2 is the default version when using REST API with Swagger UI.

**Flink****Summary:**

Remove global JobResultStore storage path

Previous behavior:

The JobResultStore was configured as a global HDFS path.

New behavior:

JobResultStore is automatically provided by YARN.

## Unsupported features

The following features are not supported in Cloudera Streaming Analytics 1.11.0.

**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 1.11.0	Apache Flink 1.16.2
CSA 1.10.0	Apache Flink 1.16.1
CSA 1.9.0	Apache Flink 1.15.1
CSA 1.8.0	
CSA 1.7.0	Apache Flink 1.14
CSA 1.6.2	
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	1.16.2
<b>CDP Private Cloud Base component support</b>	
Cloudera Runtime	7.1.9 7.1.9 SP1
Atlas	3.0.0
HBase	2.4.17
HDFS	3.1.1
Hive	3.1.3
Kafka <sup>1</sup>	3.4.1
Kudu	1.17.0
Schema Registry	0.10.0
Streams Messaging Manager	2.3.0

<b>Connector support</b>	
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
Iceberg	1.3.0

## 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

<sup>1</sup> Connecting to Kafka that is running on remote CDH6 or HDP3 is also supported.

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>1.16.2-csa1.11.0.1</version>
</dependency>
```

### Confluent Registry

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-confluent-registry</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>

<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-avro-confluent-registry</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>
```

### CSV

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-csv</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>
```

### Hive

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-hive_2.12</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>
```

### HBase

### Iceberg

```
<dependency>
<groupId>org.apache.iceberg</groupId>
<artifactId>iceberg-flink-runtime-1.16</artifactId>
<version>1.3.0.7.1.9.0-338</version>
</dependency>
```

### JSON

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-json</artifactId>
<version>1.16.2-csa1.11.0.1</version>
```

```
</dependency>
```

**Kafka**

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-kafka</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>
```

**Kudu**

```
<dependency>
<groupId>org.apache.bahir</groupId>
<artifactId>flink-connector-kudu_2.12</artifactId>
<version>1.1.0-csa1.11.0.1</version>
</dependency>
```

**Schema Registry**

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-cloudera-registry</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>
```

**Table API**

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-table-api-java-bridge</artifactId>
<version>1.16.2-csa1.11.0.1</version>
</dependency>

<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-table-planner_2.12</artifactId>
<version>1.16.2-csa1.11.0.1</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"> <li>DataStream API</li> </ul>	<ul style="list-style-type: none"> <li>ProcessFunction</li> <li>Stream Join</li> <li>Interval Join</li> <li>Stateful operators</li> <li>FsStatebackend with HDFS</li> <li>RocksDBStateBackend with HDFS</li> </ul>



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