

Cloudera Streaming Analytics 1.14.0

Release Notes

Date published: 2019-12-17

Date modified: 2024-12-03

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
Deprecation notices in Cloudera Streaming Analytics 1.14.0.....	5
Deprecation Notices for Streaming SQL Builder.....	5
Known issues and limitations.....	6
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.....	11
Flink API Support.....	13

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.14.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

Cloudera platform support

Cloudera Streaming Analytics 1.14.0 is supported as a Long Term Support (LTS) version on Cloudera Private Cloud 7.3.1. Ensure that you review the [7.3.1 Release Notes](#) and [Support Matrix](#) to understand which operating system, database, and JDK versions are supported for Streaming Analytics as well.

KNOX for SSB as Load-Balancer

From Cloudera Stream Analytics 1.14.0, SQL Stream Builder uses KNOX as a load-balancer if there are multiple instances.

The Load-balancer role, associated Cloudera Manager configurations, and nginx binaries have been removed.



Important: KNOX is the only load-balancing solution available for SQL Stream Builder. If your cluster does not support KNOX, load-balancing will not be available for SQL Stream Builder.

For more information, see [Enabling High Availability for SSB](#).



Note: KNOX requires Server-Sent Events (SSE) to be enabled in Cloudera Manager.

SQL Stream Builder sampling now using the Server-Sent Events (SSE) protocol

From Cloudera Stream Analytics 1.14.0, the sampling functionality in SQL Stream Builder is using the Server-Sent Events (SSE) protocol to support using KNOX as a load-balancer.

New Cloudera Manager configurations and REST API endpoints were added for SSE sampling.



Important: SSE support is disabled by default. To use it, you have to enable it in Cloudera Manager, see [Enabling High Availability for SSB](#).

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.14.0.

CSA-5423 - Extend SSB diag bundle data points

CSA-5440 - Permit Spring Flyway plugin execution on PvC

CSA-5364 - Add number of topics/tables to successful data source validation message on UI

CSA-5306 - SSB API does not validate catalog type

CSA-5362 - Update "ssb-sse" ASCII text banner to not contain special characters

CSA-5359 - Improve error message when creating a JS UDF with a Java version that doesn't support it

CSA-5296 - Samples table fields are limited to 32 characters in mysql and oracle dbs

CSA-5324 - SSB default admin does not have admin privileges

CSA-5428 - Sampling renders null as Invalid Number in some cases

CSA-5474 - SSB can't execute any jobs due to permission issue in the SSB artifacts directory

CSA-5475 - Local-kafka connector template not showing in SSB

CSA-5479 - Using Temp View based on kudu lookup table leaks eventpolls

CSA-5499 - Bump Avro to 1.11.4 in parcel to mitigate CVE-2024-47561

Deprecation notices in Cloudera Streaming Analytics 1.14.0

Certain features and functionalities have been removed or deprecated in Cloudera Streaming Analytics 1.14.0. You must review these items to understand whether you must modify your existing configuration. You can also learn about the features that will be removed or deprecated in the future release to plan for the required changes.

Terminology

Items in this section are designated as follows:

Deprecated

Technology that Cloudera is removing in a future Cloudera Streaming Analytics release. Marking an item as deprecated gives you time to plan for removal in a future Cloudera Streaming Analytics release.

Moving

Technology that Cloudera is moving from a future Cloudera Streaming Analytics release and is making available through an alternative Cloudera offering or subscription. Marking an item as moving gives you time to plan for removal in a future Cloudera Streaming Analytics release and plan for the alternative Cloudera offering or subscription for the technology.

Removed

Technology that Cloudera has removed from Cloudera Streaming Analytics and is no longer available or supported as of this release. Take note of technology marked as removed since it can potentially affect your upgrade plans.

Deprecation Notices for Streaming SQL Builder

Certain features and functionality are deprecated or removed in Streaming SQL Builder (SSB). You must review these changes along with the information about the features in Streaming SQL Builder (SSB) that will be removed or deprecated in a future release.

Deprecated

v1 REST API

The v1 REST API for SQL Stream Builder has been deprecated and will be removed in a future version of Streaming Analytics.

Customers are advised to migrate to the v2 API, available for SQL Stream Builder.

For more information on the v2 API, see [SQL Stream Builder REST API reference](#).

Support for JavaScript UDFs

Due to the deprecation of the Nashorn engine used in JDK 8 and 11, User-Defined Functions (UDFs) written in JavaScript has been deprecated in Cloudera Streaming Analytics 1.13.0. Cloudera recommends that customers start using [Python UDFs](#) for all new developments, and start migrating their JavaScript UDFs to Python to prepare for future upgrades.

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 1.14.0.

SQL Stream Builder

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_config IS NULL;
```

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.

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](#).

Flink

CSA-5525 - Illegal join reordering in Flink optimizer

Flink optimizer's reordering might violate certain clauses (for example FOR SYSTEM_TIME AS OF) that are supported only on a specific side of a join operation, resulting in an error.

Example error message:

```
Caused by: org.apache.flink.table.api.TableException: Temporal table join only support apply FOR SYSTEM_TIME AS OF on the right table
```

Set table.optimizer.join-reorder-enabled to false, until the issue is fixed in upstream Flink.

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.

Summary:

Use KNOX for SQL Stream Builder as Load-Balancer

Previous behavior:

SQL Stream Builder used a Load Balancer instance with nginx to support high availability.

New behavior:

SQL Stream Builder only uses KNOX as a load-balancer if there are multiple instances. See [Enabling High Availability for SSB](#).

Summary:

SQL Stream Builder sampling now uses the Server-Sent Events (SSE) protocol.

Previous behavior:

SQL Stream builder samples were delivered via a websocket connection.

New behavior:

The sampling functionality in SQL Stream Builder is using the Server-Sent Events (SSE) protocol to support using KNOX as a load-balancer. See [Enabling High Availability for SSB](#).

Unsupported features

Some Apache Flink and SSB features exist in Cloudera Streaming Analytics 1.14.0, but are not supported by Cloudera. These features are not ready for production deployment, but Cloudera encourages you to explore them in non-production environments and provide feedback on your experiences through the Cloudera Community Forums.

SQL Stream Builder

- Virtual environments for Python are not supported.
- 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
- 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 1.14.0	Apache Flink 1.19.1
CSA 1.13.0	Apache Flink 1.19.1
CSA 1.12.0	Apache Flink 1.18.0
CSA 1.11.0	Apache Flink 1.16.2
CSA 1.10	Apache Flink 1.16.1
CSA 1.9.0	Apache Flink 1.15.1
CSA 1.8.0	
CSA 1.7.0	
CSA 1.6.2	Apache Flink 1.14
CSA 1.6.1	

CSA version	Component version
CSA 1.6.0	
CSA 1.5.3	
CSA 1.5.1	Apache Flink 1.13
CSA 1.5.0	
CSA 1.4.1	
CSA 1.4.0	Apache Flink 1.12
CSA 1.3.0	
CSA 1.2.0	Apache Flink 1.10
CSA 1.1.0	Apache Flink 1.9.1

Related Information

[CSA 1.14.0 Release Notes](#)

[CSA 1.13.2 Release Notes](#)

[CSA 1.13.1 Release Notes](#)

[CSA 1.13.0 Release Notes](#)

[CSA 1.12.0 Release Notes](#)

[CSA 1.11.0 Release Notes](#)

[CSA 1.10.0 Release Notes](#)

[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	1.19.1
Cloudera Runtime component support in Cloudera Private Cloud Base Cloudera 7.3.1	
Atlas	2.1.0
HBase	2.4.17
HDFS	3.1.1

Apache Flink support	1.19.1
Cloudera Runtime component support in Cloudera Private Cloud Base Cloudera 7.3.1	
Hive	3.1.3
Kafka ¹	3.4.1
Kudu	1.17.0
Schema Registry	0.10.0
Streams Messaging Manager	2.3.0
Apache Iceberg	1.3.1

Connector support	
JDBC PostgreSQL	9.6-16
JDBC MySQL	5.7, 8
JDBC Hive	3.1.3
JDBC Oracle	19, 19c, 21c, 23c
JDBC Db2	11.5
JDBC SQL Server	2007-2022
CDC PostgreSQL	9.6-16
CDC MySQL	5.7, 8
CDC Oracle	19, 19c, 21c, 23c
CDC Db2	11.5
CDC SQL Server	2007-2022
Apache Iceberg	1.3.1

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

¹ 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>1.19.1-csa1.14.0.0</version>
</dependency>
```

Confluent Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-confluent-registry</artifactId>
  <version>1.0-csa1.14.0.0</version>
</dependency>
```

CSV

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-csv</artifactId>
  <version>1.19.1-csa1.14.0.0</version>
</dependency>
```

Hive

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hive_2.12</artifactId>
  <version>1.19.1-csa1.14.0.0</version>
</dependency>
```

HBase

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-1.4</artifactId>
  <version>3.0-csa1.14.0.0</version>
</dependency>
```

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-2.4</artifactId>
  <version>3.0-csa1.14.0.0</version>
</dependency>
```

Iceberg

```
<dependency>
  <groupId>org.apache.iceberg</groupId>
  <artifactId>iceberg-flink-runtime-1.17</artifactId>
  <version>1.3.1.7.3.1.0-197</version>
```

```
</dependency>
```

JDBC

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-jdbc</artifactId>
  <version>3.2-csa1.14.0.0</version>
</dependency>
```

JSON

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-json</artifactId>
  <version>1.19.1-csa1.14.0.0</version>
</dependency>
```

Kafka

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-kafka</artifactId>
  <version>3.2-csa1.14.0.0</version>
</dependency>
```

Kudu

```
<dependency>
  <groupId>org.apache.bahir</groupId>
  <artifactId>flink-connector-kudu_2.12</artifactId>
  <version>1.1.0-csa1.14.0.0</version>
</dependency>
```

Schema Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-cloudera-registry</artifactId>
  <version>1.0-csa1.14.0.0</version>
</dependency>
```

Table API

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-api-java-bridge</artifactId>
  <version>1.19.1-csa1.14.0.0</version>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-planner-loader-bundle</artifactId>
  <version>1.19.1-csa1.14.0.0</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">ProcessFunctionStream JoinInterval JoinStateful operatorsFsStatebackend with HDFSRocksDBStateBackend with HDFS



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