

Release Notes

Date published: 2019-12-17

Date modified: 2023-02-23



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

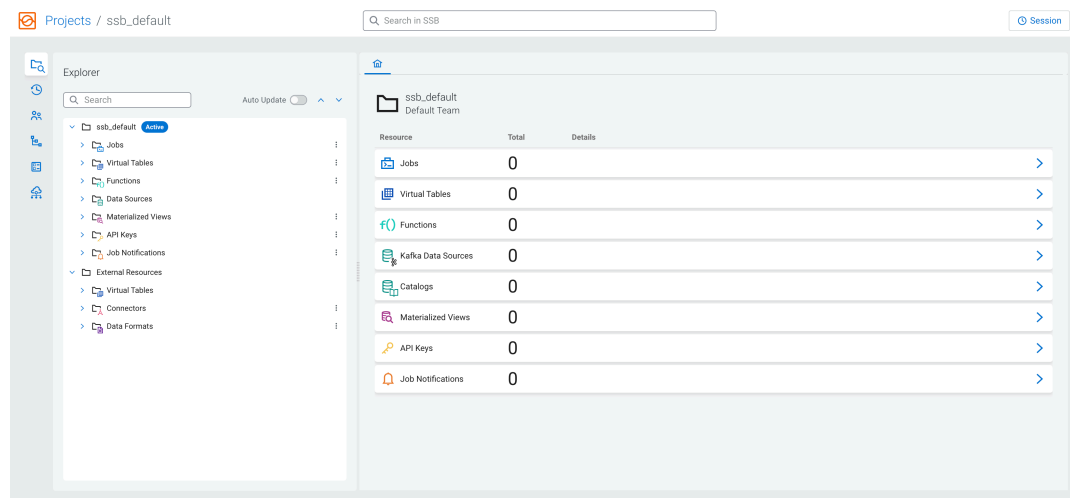
What's new in Cloudera Streaming Analytics.....	4
Fixed issues.....	4
Known issues and limitations.....	5
Behavioral changes.....	8
Unsupported features.....	8
Support Matrix.....	9
Component support.....	9
System Requirements.....	10
Default ports for Flink and SSB.....	10
Maven dependencies in Flink.....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

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.9.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

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. For more information, see the [Project structure and development](#) documentation.



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](#).

Confluent Schema Registry support

Confluent Schema Registry can be used as a catalog in SQL Stream Builder and Flink. For more information, see the [Adding Catalogs](#) documentation.

JSON support for Schema Registry

JSON schemas are supported when using Cloudera Schema Registry with Kafka.



Note: JSON support is available when using CSA on CDP Private Cloud Base 7.1.8.

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.9.0.

CSA-2016: Deleting table from other teams

The issue about deleting table from other teams is fixed.

CSA-3867 - UDF case sensitivity

The issue regarding case sensitivity for UDFs is fixed.

CSA-3956 - SSB fails to parse SQL with comments

The issue regarding the failure to parse SQL queries with comments is fixed.

CSA-4044 - SSB keytab upload fails if keytab working dir is deleted

The issue regarding the keytab directory being deleted from the tmp folder is fixed.

CSA-4363 - Production mode does not use clean session variables

The issue about production mode not using clean session variables 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 1.9.0.

SQL Stream Builder

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

When querying data from a table or a view with a ROW() function an exception is thrown due to a Calcite parsing issue. For example, the following query will return an error:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM table;
SELECT * FROM example;
```

Add a second SELECT layer to the SQL query as shown in the following example:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM (SELECT col1,
col2 FROM table);
SELECT * FROM example;
```

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.

Streaming SQL Console cannot be accessed through Knox when High Availability is enabled

When SQL Stream Builder SSB is deployed in High Availability with Load Balancer, the Streaming SQL Console cannot be accessed directly using Apache Knox.

For accessing Streaming SQL Console, use the secured Load Balancer deployment or authenticate using SPNEGO.

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.

CSA-4650: Inconsistent sidebar collapse behavior

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

None

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-4548: Files cannot be uploaded through Swagger

REST API endpoints that take multipart requests, such as the uploading artifact endpoint, result in error.

None

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

When a job is opened in **Tab** view, the Execute and Stop actions are disabled when managing the job from the **Explorer** view.

None

CSA-4426: Kafka Data Source name accepts spaces

Kafka Data Source can be validated and created with spaces in the Data Source name, but this results in errors as spaces are not valid characters based on the naming convention.

None

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

The show password icon can be used after saving the password for the authentication method when creating a Kafka Data Source.

None

CSA-4412: Cannot delete Materialized View endpoint when using dynamic parameters

Materialized View endpoints cannot be deleted if dynamic parameters were set for them.

None

CSA-4400: Cannot delete invalid catalog

A catalog can be created without a catalog service, but deleting the invalid catalog fails as it is not registered in Flink without a service.

None

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

Describing fails for a Virtual Table from a schema in a Schema Registry Catalog when viewing the table DDL.

None

CSA-4333: Use Kafka Timestamps switch reflects invalid value

After creating a Kafka Virtual Table and disabling the Use Kafka Timestamps configuration, the table is created successfully according to the setting, but when viewing the DDL of the table, it shows the configuration as enabled.

None

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

When creating a webhook table with a custom template, the webhook template will be saved as an empty string, which results in webhook sending failure.

None

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

The loadbalancer.url property is duplicated in Cloudera Manager on the SQL Stream Builder configuration page.

None

Flink**FLINK-18027: ROW value constructor cannot deal with complex expressions**

When querying data from a table or a view with a ROW() function an exception is thrown due to a Calcite parsing issue. For example, the following query will return an error:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM table;
SELECT * FROM example;
```

Add a second SELECT layer to the SQL query as shown in the following example:

```
CREATE VIEW example AS SELECT col1, ROW(col2) FROM (SELECT col1,
col2 FROM table);
SELECT * FROM example;
```

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:

Some public API endpoints got reworked and have different endpoints or related models because of some business logic decisions.

Previous behavior:

For more information about the previous behavior, see the [previous version of the SQL Stream Builder API Reference documentation](#).

New behavior:

For more information about the new behavior, see the [latest version of the SQL Stream Builder API Reference documentation](#).

Flink

Summary:

Additional step required when upgrading from a lower version of CSA 1.8.0 to the 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 1.9.0.

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
- Python API

- 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 1.9.0	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 1.8.0 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	1.15.1	
CDP Private Cloud Base component support		
Cloudera Runtime	7.1.7 SP2 7.1.7 SP3	7.1.8
Kafka ¹	2.5.0	3.1.1
Schema Registry	0.10.0	0.10.0
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 port 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

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

Component	Service	Port	Configuration property
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>1.15.1-cs1.9.0.0</version>
</dependency>
```

Confluent Registry

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-confluent-registry</artifactId>
<version>1.15.1-cs1.9.0.0</version>
</dependency>

<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-avro-confluent-registry</artifactId>
<version>1.15.1-cs1.9.0.0</version>
</dependency>
```

CSV

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-csv</artifactId>
<version>1.15.1-cs1.9.0.0</version>
</dependency>
```

Hive

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-hive_2.12</artifactId>
<version>1.15.1-cs1.9.0.0</version>
</dependency>
```

HBase

Iceberg

JSON

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-json</artifactId>
<version>1.15.1-csa1.9.0.0</version>
</dependency>
```

Kafka

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-connector-kafka</artifactId>
<version>1.15.1-csa1.9.0.0</version>
</dependency>
```

Kudu

```
<dependency>
<groupId>org.apache.bahir</groupId>
<artifactId>flink-connector-kudu_2.12</artifactId>
<version>1.1.0-csa1.9.0.0</version>
</dependency>
```

Schema Registry

```
<dependency>
<groupId>org.apache.flink</groupId>
<artifactId>flink-cloudera-registry</artifactId>
<version>1.15.1-csa1.9.0.0</version>
</dependency>
```

Table API

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-api-java-bridge</artifactId>
  <version>1.15.1-csa1.9.0.0</version>
</dependency>

<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-planner_2.12</artifactId>
  <version>1.15.1-csa1.9.0.0</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).