

Cloudera Streaming Analytics 1.12.0

Release Notes

Date published: 2019-12-17

Date modified: 2024-02-12

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.....	4
Behavioral changes.....	6
Unsupported features.....	7
Support Matrix.....	7
Component support.....	7
System Requirements.....	8
Default ports for Flink and SSB.....	9
Maven dependencies in Flink.....	10
Flink API Support.....	11

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.12.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

Rebase to Apache Flink 1.18.0

Apache Flink 1.18.0 is supported in Cloudera Streaming Analytics 1.12.0.

For more information on what is included in the Apache Flink 1.18.0 version, see the [Apache Flink 1.18.0 Release Announcement](#) and [Release Notes](#).

Application Mode

Support for Flink's Application Mode is added. Application Mode enables you to submit Flink applications that run entirely on the cluster side, differently from Per-Job and Session modes where the application driver code runs on the client side. In Application Mode, each application can execute multiple jobs that share cluster resources but are isolated from other applications.

For more information, see the [Running a Flink job](#) documentation.

Support for custom port range in Task Manager

Support is added to configure `taskmanager.data.port` in Flink service configuration. This property allows for the configuration of custom port ranges for Flink jobs' Task Manager processes. The `taskmanager.data.port` property can be configured for Flink in Cloudera Manager using a safety valve, and for SSB the property can be configured with SET statement.

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.12.0.

- CSA-4973 - If auto TrustStore is enabled, validation fails on existing Kafka sources
- CSA-4981 - Disable widget auto update when authentication error occurs for Materialized View data source
- CSA-4988 - Restrict connector and data format uploads to administrators
- CSA-5006 - SSB service fails when using Active Directory (AD) Kerberos authentication

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 1.12.0.

CSA-4464: CSA parcel is built with an interim CDP build

The CSA parcel is built using an interim CDP build and not with a build that corresponds to a release version. This can cause errors with Flink components that have dependency to CDP.

If a Flink component transitively depends on some CDP related module and it is not accessible publicly, the dependency can be excluded. In case the project also depends on the excluded module, the publicly available version of the dependency can be added to Flink. For example, this could happen with `kafka-clients`, which is pulled in by `flink-connector-kafka`:

```
...
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-kafka</artifactId>
  <version>${flink.version}</version>
  <exclusions>
    <exclusion>
```

```

    <groupId>org.apache.kafka</groupId>
    <artifactId>kafka-clients</artifactId>
  </exclusion>
</exclusions>
</dependency>
<dependency>
  <groupId>org.apache.kafka</groupId>
  <artifactId>kafka-clients</artifactId>
  <version>3.1.2.7.2.16.0-287</version>
</dependency>
...

```

For more information about the supported dependency version, see the [Maven dependencies in Flink](#) documentation.

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 compiled.

```

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

None

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.

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.

Summary:

The user-defined function (UDF) tester is not available on Streaming SQL Console

Previous behavior:

New UDFs could be tested on Streaming SQL Console before executing the SQL job.

New behavior:

The UDF test runner is turned off, the added UDFs can be tested with executing the SQL job.

Summary:

Default state backend is changed to RocksDB

Previous behavior:

Hashmap was configured as the default state backend.

New behavior:

RocksDB is configured as the default state backend.

Unsupported features

The following features are not supported in Cloudera Streaming Analytics 1.12.0.

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
- 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 1.12.0	Apache Flink 1.18.0

CSA version	Component version
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	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 1.11.0 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	1.18.0
Cloudera Runtime component support in 7.1.9	
Atlas	3.0.0
HBase	2.4.17
HDFS	3.1.1
Hive	3.1.3
Kafka ¹	3.4.1
Kudu	1.17.0
Schema Registry	0.10.0
Streams Messaging Manager	2.3.0

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
Apache 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

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

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

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>1.18.0-csa1.12.0.0</version>
</dependency>
```

Confluent Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-confluent-registry</artifactId>
  <version>1.0-csa1.12.0.0</version>
</dependency>
```

CSV

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-csv</artifactId>
  <version>1.18.0-csa1.12.0.0</version>
</dependency>
```

Hive

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hive_2.12</artifactId>
  <version>1.18.0-csa1.12.0.0</version>
</dependency>
```

HBase

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-1.4</artifactId>
  <version>3.0-csa1.12.0.0</version>
</dependency>
```

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-2.4</artifactId>
  <version>3.0-csa1.12.0.0</version>
</dependency>
```

Iceberg

```
<dependency>
  <groupId>org.apache.iceberg</groupId>
  <artifactId>iceberg-flink-runtime-1.16</artifactId>
  <version>1.3.0.7.1.9.1-158</version>
</dependency>
```

JSON

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-json</artifactId>
  <version>1.18.0-csa1.12.0.0</version>
</dependency>
```

Kafka

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-kafka</artifactId>
  <version>3.1-csa1.12.0.0</version>
</dependency>
```

Kudu

```
<dependency>
  <groupId>org.apache.bahir</groupId>
  <artifactId>flink-connector-kudu_2.12</artifactId>
  <version>1.1.0-csa1.12.0.0</version>
</dependency>
```

Schema Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-cloudera-registry</artifactId>
  <version>1.0-csa1.12.0.0</version>
</dependency>
```

Table API

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-api-java-bridge</artifactId>
  <version>1.18.0-csa1.12.0.0</version>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-planner_2.12</artifactId>
  <version>1.18.0-csa1.12.0.0</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).