Cloudera Streaming Analytics 1.13.0

Release Notes

Date published: 2019-12-17
Date modified: 2024-08-07

https://docs.cloudera.com/

Legal Notice

Contents

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.13.0 covers new features beside the core streaming functionality of Apache Flink and SQL Stream Builder.

Rebase to Apache Flink 1.19.1

Apache Flink 1.19.1 is supported in Cloudera Streaming Analytics 1.13.0.

For more information on what is included in the Apache Flink 1.19.1 version, see the Apache Flink 1.19.1 Release Announcement and Release Notes.

Support for Python UDFs in SSB

This feature allows customers to start using Python for creating User-Defined Functions (UDFs). 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, as Javascript UDFs will be removed in the future due to the deprecation of the Nashorn engine used in JDK 8 and 11.

For more information on using Python UDFs, see Python UDFs. For more information on supported JDK versions, refer to the Support Matrix.

Global logging configuration for SSB jobs

A new global settings view has been enabled, which currently includes log4j configuration of Flink jobs started on SSB. Users with SSB administrator rights can set a default logging configuration applied to all SSB jobs , which can be overridden at the job level.

For more information see Adjusting logging configuration in Advanced Settings.

Customizable default Kafka TrustStore configuration in Streaming SQL Console

Customizing default Kafka TrustStore configurations was added to Streaming SQL Console. Kafka TrustStore can be configured during adding Kafka as a Data Source on the UI.

Cloudera platform support

Cloudera Streaming Analytics 1.13.0 is supported as a Long Term Support (LTS) version on CDP Private Cloud Base 7.1.9 SP1. Ensure that you review the 7.1.9 SP1 Release Notes and Support Matrix to understand which operating system, database, and JDK versions are supported for Streaming Analytics as well.

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.13.0.

Cannot submit SQL jobs with UDF JARs when checkpoints enabled

CSA-5048 - Generate correct default log configuration for SSB jobs

CSA-5055 - Backport FLINK-20539 to CSA

CSA-5065 - Artifact Storage request thread does not timeout when storage is offline, hanging the UI

CSA-5120 - Connector dependencies are missing from SSB

CSA-5122 - SSB keeps reconciling after a failed job

CSA-5126 - Undeterministic classloader behavior with flink-metrics-kafka and kafka-connector resulting in job failure

CSA-5161 - analyzeQuery returning false validation errors in some cases

CSA-5166 - SSB Local Kafka data source doesn't work with a user-specified TLS truststore

CSA-5199 - Cloudera registry catalog type registry is not compatible with UI

CSA-5221 - SsbCatalog uses user session in prod mode

CSA-5236 - Implement MetricReporterFactory for KafkaMetricsReporter

CSA-5251 - UiConfigController throws NullPointerException

CSA-5270 - If custom log is set, the job cannot be saved

CSA-5282 - Check Flink job status before submitting SSB job

CSA-5283 - [ssb] Make all overloaded methods transactional in JobService

CSA-5303 - Fix Kerberos/SPNEGO authentication for Flink Deployments

CSA-5306 - SSB API does not validate catalog type

CSA-5315 - Load balancer role cannot start

UI fixes and improvements

CSA-4602 - Changing existing MV filter type can't be saved

CSA-5038 - Widget is empty when added to the dashboard before initialization completes

CSA-5140 - Fix No Rows To Show message when switching from sampler to MV in dashboard preview

CSA-5024 - Polling samples feedback is on even when polling is turned off

CSA-5025 - Cursor jumps to the end after the first keystroke when using templates

CSA-5026 - Oversize widget cannot be sized down

CSA-5294 - Add job save button to job settings component

Deprecation notices in Cloudera Streaming Analytics 1.13

Certain features and functionalities have been removed or deprecated in Cloudera Streaming Analytics 1.13. 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

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 are deprecated in Cloudera Streaming Analytics 1.13. 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 1.13.0.

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_conf
ig 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

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:

CSA-5068 - Page refresh is not required to update result samples.

Previous behavior:

Sample IDs could become invalid if another user restarted the job while polling in the UI.

New behavior:

Sample IDs are updated before and during polling.

Summary:

CSA-5238 - YARN is set as default token renewer automatically based on configuration.

Previous behavior:

Configuration check wasn't performed and YARN wasn't set as the default renewer.

New behavior:

YARN is set as the default token renewer if the deployment target is yarn-session or yarn-per-job.

Summary:

CSA-5258 - Extended validation for dynamic MV parameter names.

Previous behavior:

Dynamic pattern validation was restrictive which could result in errors when querying MV.

New behavior:

Validation pattern is extended to [A-Za-z_\-0-9\.\~]+

Summary:

CSA-5214 - Add option to SSB service to customize default truststore.

Previous behavior:

Default Kafka TrustStore configurations could only be customized in Flink configurations.

New behavior:

Customizing default Kafka TrustStore configurations was added to Streaming SQL Console. Kafka TrustStore can be configured during adding Kafka as a Data Source on the UI.

Summary:

CSA-5199 - API change: the Cloudera registry catalog type registry is now named cloudera-reg istry.

Previous behavior:

When creating a Schema Registry data source catalog type property had to be registry.

New behavior:

Catalog type property now has to be cloudera-registry when creating a Schema Registry data source.

Note: Using registry is backward compatible.

Summary:

CSA-5306 - SSB API does not validate data sources before saving.

Previous behavior:

There was no validation before saving a data source via the API. If the user wanted to make sure to save a valid data source they had to use the validate endpoint before saving. When creating a data source on SSB UI, the user had the option to save an invalid data source.

New behavior:

Saving the data source is only allowed if the data source is valid. SSB API validates the data source before saving and on the SSB UI Create / Save button is only active if the validation is successful.

Unsupported features

Some Apache Flink and SSB features exist in Cloudera Streaming Analytics 1.13.0, 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 1.13.0	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	Apache Flink 1.14
CSA 1.6.2	
CSA 1.6.1	
CSA 1.6.0	

CSA version	Component version
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 1.13.0 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 CDP Private Cloud Base CDP 7.1.9 SP1	
Atlas	3.0.0
HBase	2.4.17
HDFS	3.1.1
Hive	3.1.3
Kafka[1]	3.4.1

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

Apache Flink support	1.19.1
Cloudera Runtime component support in CDP Private Cloud Base CDP 7.1.9 SP1	
Kudu	1.17.0
Schema Registry	0.10.0
Streams Messaging Manager	2.3.0
Apache Iceberg	1.3.0

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	1.3.0

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 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
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>1.19.1-csa1.13.0.0</version>
</dependency>
```

Confluent Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-confluent-registry</artifactId>
  <version>1.0-csa1.13.0.0</version>
</dependency>
```

CSV

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-csv</artifactId>
  <version>1.19.1-csa1.13.0.0</version>
</dependency>
```

Hive

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hive_2.12</artifactId>
  <version>1.19.1-csa1.13.0.0</version>
</dependency>
```

HBase

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-1.4</artifactId>
  <version>3.0-csa1.13.0.0</version>
</dependency>
```

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-hbase-2.4</artifactId>
  <version>3.0-csa1.13.0.0</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>
```

JDBC

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-jdbc</artifactId>
  <version>3.2-csa1.13.0.0</version>
</dependency>
```

JSON

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-json</artifactId>
  <version>1.19.1-csa1.13.0.0</version>
</dependency>
```

Kafka

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-kafka</artifactId>
  <version>3.2-csa1.13.0.0</version>
</dependency>
```

Kudu

```
<dependency>
  <groupId>org.apache.bahir</groupId>
  <artifactId>flink-connector-kudu_2.12</artifactId>
  <version>1.1.0-csa1.13.0.0</version>
</dependency>
```

Schema Registry

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-connector-cloudera-registry</artifactId>
  <version>1.0-csa1.13.0.0</version>
</dependency>
```

Table API

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-api-java-bridge</artifactId>
  <version>1.19.1-csa1.13.0.0</version>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-planner_2.12</artifactId>
  <version>1.19.1-csa1.13.0.0</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)
• DataStream API	• ProcessFunction • Stream Join • Interval Join • Stateful operators • FsStatebackend with HDFS • RocksDBStateBackend with HDFS

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