Cloudera Streaming Analytics 1.5.0

Release Notes

Date published: 2019-12-17
Date modified: 2021-09-29

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

CVE-2021-45105 & CVE-2021-44832 remediation for CSA

Learn more about the CVE-2021-45105 and CVE-2021-44832 remediation for Flink and SQL Stream Builder in Cloudera Streaming Analytics (CSA).

Cloudera released maintenance versions for CSA on CDP Private Cloud Base to address CVE-2021-45105 and CVE-2021-44832 identified as critical vulnerability issues for Log4j2.

Cloudera encourages users to upgrade to the following CSA versions to avoid any possibility of exploitation:

- CVE-2021-45105: CSA 1.5.3 or higher version
- CVE-2021-44832: CSA 1.6.1

For more information about the impacts of CVE-2021-45105, see the TSB 2021-547: Critical vulnerability in log4j2 CVE-2021-45105 Knowledge Base article.

Related Information

CSA 1.5.3 Installation guide

CSA 1.6.1 Installation guide

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.5.0 covers new features beside the core streaming functionality of Apache Flink.

Cloudera Streaming Analytics 1.5.0 introduces the following key features:

Apache Flink upgrade

Apache Flink 1.13 is supported in Cloudera Streaming Analytics 1.5.0.

SQL Session Cluster support

Session mode is supported for SQL Stream Builder which means that the Flink jobs are running in a session cluster by default. You are also able to set the properties for the session mode directly from the SQL window on the Streaming SQL Console.

For more information, see the Managing session for SQL jobs section.

Connector support

The following new connectors are added as Templates to SQL Stream Builder:

- Change Data Capture (CDC) connectors: Oracle, MySQL, PostgreSQL
- Blackhole connector
- Faker connector

You are also able to use more types of Kafka connectors as there are individual templates for Upsert Kafka, Kafka in CDP and Kafka outside of CDP.

For more information, see the Supported tables in SQL Stream Builder section.

REST API support

You can use the SQL Stream Builder REST API to manage and monitor your SQL jobs from the CLI using POST and GET endpoints.

For more information, see the Using SQL Stream Builder REST API section and the SQL Stream Builder API Reference documentation.

Knox Auto Discovery

The Auto Discovery feature of Knox is supported for Flink and SQL Stream Builder. This means that you do not need to manually configure the topologies and service definitions for the Knox integration.

For more information, see the Enabling Knox for SQL Stream Builder and Enabling Knox for Flink Dashboard sections.

Streaming SQL Console updates

Expanded Flink SQL support

More SQL statements are available that can be used directly from the SQL window on the Streaming SQL Console.

INSERT INTO support

Instead of using a Sink Table on the Streaming SQL Console, you are able to use the INSERT INTO statement directly in the SQL window to output your query results.

Tables > Flink DDL option is no longer supported

Instead of using the Flink DDL wizard, you are able to use Templates under the SQL window to select and import the CREATE TABLE templates.

For more information, see the Concept of tables in SQL Stream Builder section.

Component support

Learn more about which Apache Flink component version is supported in the Cloudera Streaming Analytics (CSA) releases.

CSA version	Component version
CSA 1.5.0	Apache Flink 1.13
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 1.4.1 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

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 1.5.0.

SQL Stream Builder

CSA-2302: SQL Stream Builder ignores restart from savepoint

SQL Stream Builder ignores the option to restart from savepoint, and starts the SQL job without the state stored at the latest savepoint.

None

CSA-2321: SASL/PLAINTEXT is not available

The SASL/PLAINTEXT option is not available on Streaming SQL Console when adding Kafka as a Data Provider.

None

CSA-2301: Transformation error when Kafka has null values

An error occurs when using the Input transform on Kafka messages in raw format without values.

None

CSA-2156: SASL password is displayed for Kafka Data Provider

The password you set for the Kafka Data Provider when using the wizard in Streaming SQL Console, displays the password value to the UI.

None

CSA-2155: Dependency version mismatch in Streaming SQL Engine logs

Errors can occur in the Streaming SQL Engine logs as the JAR files on the Streaming SQL Engine classpath can have different versions. This causes a dependency version issue in Streaming SQL Engine.

None

CSA-2093: SQL Jobs are not listed with Knox

When using Knox authentication, the stopped or running jobs are not listed under the SQL Jobs tab in Streaming SQL Console.

None

CSA-2016: Deleting table from other teams

There is a limitation when using the Streaming SQL Console for deleting tables. It is not possible to delete a table that belongs to another team using the Delete button on the User Interface.

Use DROP TABLE statement from the SQL window.

CSA-1985: DROP TABLE limitation when using Webhook table

DROP TABLE cannot be executed against Webhook type tables. The following error message is displayed when trying to delete a Webhook table using the SQL window: Table with identifier 'xyz' does not exist.

Use the Delete button on the Streaming SQL Console.

CSA-1673: SSB operations are not showing in Atlas

Due to a communication issue SQL Stream Builder (SSB) operations are not showing in Atlas.

None

CSA-1454: Timezone settings can cause unexpected behavior in Kafka tables

You must consider the timezone settings of your environment when using timestamps in a Kafka table as it can affect the results of your query. When the timestamp in a query is identified with from_unixtime, it returns the results based on the timezone of the system. If the timezone is not set in UTC+0, the timestamp of the query results will shift in time and will not be correct.

Change your local timezone settings to UTC+0.

CSA-1231: Big numbers are incorrectly represented on the Streaming SQL Console UI

The issue impacts the following scenarios in Streaming SQL Console:

- When having integers bigger than 253-1 among your values, the Input transformations and User Defined Functions are considered unsafe and produce incorrect results as these numbers will lose precision during parsing.
- When having integers bigger than 253-1 among your values, sampling to the Streaming SQL Console UI produces incorrect results as these numbers will lose precision during parsing.

None

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.

Unsupported features

The following features are not supported in Cloudera Streaming Analytics 1.5.0.

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