Cloudera Streaming Analytics 1.7.0

Release Notes

Date published: 2019-12-17
Date modified: 2022-05-05

https://docs.cloudera.com/

Legal Notice

Contents

What's new in Cloudera Streaming Analytics

Cloudera Streaming Analytics 1.7.0 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.

High Availability for SSB

You can use SQL Stream Builder (SSB) with a Load Balancer to distribute tasks over resources in case of a single point of failure.

SQL Server CDC Connector support

MS SQL Server Change Data Capture (CDC) connector is added to the list of supported connectors. The SQL Server CDC connector can be used with Flink DDL on the Streaming SQL Console either directly from the SQL Editor or using the DDL template.

Configuring retention time for Materialized Views

You can configure how data should be retained for Materialized Views based on time or row count.

Fixed issues

Review the list of Flink and SQL Stream Builder issues that are resolved in Cloudera Streaming Analytics 1.7.0.

SQL Stream Builder

CSA-3308: Changing Primary Key for Materialized View without recreating table causes job failure

The issue regarding job failure when changing the primary key for Materialized View has been fixed.

Flink

FLINK-27441: Scrollbar is missing for particular UI elements on Flink Dashboard

The issue regarding the visibility of scrollbar on t he Flink Dashboard 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 1.7.0.

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.

CSA-5006: SSB service fails when using Active Directory (AD) Kerberos

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-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

CSA-3536: SSB jobs sometimes fail to stop correctly

This issue manifests itself as a job continuing to run after clicking the Stop button on **Compose** window of the Streaming SQL Console.

You can cancel the job from Flink Dashboard in case you cannot stop it on the Streaming SQL Console:

1. Reload the window of your browser.
2. Click Flink Dashboard from the main menu.
3. Select the Cluster link for the job that failed to stop.
4. Select the job from the list.
5. Select Cancel Job at the top right of the page.

CSA-3529: Filtering ARRAY typed columns do not work for Materialized Views

ARRAY typed columns are treated as STRING when setting filters for a Materialized View query which leads to incorrect results.

None

CSA-3507: IN and NOT IN operators cannot be used for Materialized Views

Using IN and NOT IN operators when filtering Materialized Views throws an error and cannot be used.

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-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

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

Removing Streaming SQL Console instance

Due to the architectural changes of SSB, the Streaming SQL Console instance is not needed anymore for the SSB service. This means that you need to remove the Streaming SQL Console instance from your cluster when upgrading to the 1.7.0 version.

For more information, see Upgrading CSA artifacts for SQL Stream Builder service documentation.

Apache Knox authentication requires manual setup

Previous behavior:

Auto discovery of Knox authentication is supported for the following CSA versions:

- CSA 1.5.0, 1.5.1, 1.5.3 (only when using CDP Private Cloud Base 7.1.7)
- CSA 1.6.0, 1.6.1, 1.6.2, 1.6.3

New behavior:

In case you are upgrading from a CSA version where Knox auto discovery is supported, you need to manually configure Knox. The exact configuration process depends on the CDP Private Cloud Base version.

For more information, see Enabling Knox authentication for SSB documentation.

Unsupported features

The following features are not supported in Cloudera Streaming Analytics 1.7.0.

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

Component support

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

CSA version	Component version
CSA 1.7.0	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 1.6.2 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