Cloudera Streaming Analytics Operator 1.5.0

Cloudera Streaming Analytics Operator for Kubernetes 1.5.0 Release Notes

Date published: 2024-06-15
Date modified: 2026-02-18

https://docs.cloudera.com/

Legal Notice

Contents

Cloudera Streaming Analytics Operator for Kubernetes 1.5.0

Learn about the new features, improvements, known and fixed issues, limitations, and unsupported features in this release of Cloudera Streaming Analytics Operator for Kubernetes.

What's new

Learn about the new features and notable changes in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Cloudera Streaming Analytics Operator for Kubernetes 1.5.0

To learn more about Cloudera Streaming Analytics Operator for Kubernetes and its typical deployment architecture, see Overview and for the installation instructions, see Installation.

This release of Cloudera Streaming Analytics Operator for Kubernetes is based on Apache Flink Kubernetes Operator 1.13 and Apache Flink 1.20.

- Flink Kubernetes Operator 1.13 release announcement
- Flink 1.20 release announcement

Materialized View engine on Kubernetes

Cloudera SQL Stream Builder now supports the Materialized View (MV) engine on Kubernetes deployments. You can configure and create Materialized Views to maintain mutating snapshots of queried data, which can be accessed through REST APIs.

Asynchronous Flink job handling in Cloudera SQL Stream Builder

Cloudera SQL Stream Builder now handles Flink jobs asynchronously. This improves Streaming SQL Console responsiveness and stabilizes job status transitions during deployment and state changes.

Apache Flink Kubernetes Operator rebased to 1.13

The embedded Apache Flink Kubernetes Operator has been rebased to 1.13, bringing upstream stability improvements and fixes.

Default PostgreSQL version updated to 18.1

The default PostgreSQL database deployed with Cloudera SQL Stream Builder is now version 18.1.

Java runtime compatibility update

Docker images for Cloudera Streaming Analytics Operator for Kubernetes components now use an OpenJDK 17 build that supports older cipher suites for broader compatibility with legacy security configurations.

Related Information

Python UDFs in Cloudera SQL Stream Builder | Cloudera Streaming Analytics Documentation

Adjusting logging configuration in Advanced Settings | Cloudera Streaming Analytics Documentation

Fixed issues and improvements

Learn about the fixed issues, improvements, and changes functionality in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Fixed issues

CSADS-47: Cloudera SQL Stream Builder intermittently picked up transient FlinkDeployment errors

Previously, intermittent errors from the FlinkDeployment resource could be surfaced by Cloudera SQL Stream Builder, which could lead to application failure. This issue is fixed.

CSA-5960: Missing Iceberg catalog in default catalogs for Cloudera SQL Stream Builder

Cloudera SQL Stream Builder now supports Iceberg as a first-class catalog type under Data Sources and Catalogs, including Hive-backed Iceberg catalog configuration. Iceberg tables are no longer presented under Hive catalogs, preventing access errors through the Hive connector.

Improvements

Learn about the fixed issues and improvements in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Improvements

Learn about selected product improvements in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Asynchronous Flink job handling in Cloudera SQL Stream Builder

Flink jobs are now handled asynchronously, improving Streaming SQL Console responsiveness and stabilizing job state transitions during deployment and runtime changes.

Behavioral changes

Learn about the change in certain functionality of Flink and Cloudera SQL Stream Builder that has resulted in a change in behavior from the previously released version of Cloudera Streaming Analytics Operator for Kubernetes.

Behavioral changes

Summary: Updated default resource requests

The default CPU and memory resource requests in the Helm chart values.yaml are updated in this release.

During upgrade, review your values.yaml and ensure your cluster has enough capacity for the new defaults, or override these values to match your previous configuration.

Known issues and limitations

Learn about the known issues and limitations in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Stuck session jobs in Cloudera Streaming Analytics Operator for Kubernetes

Session jobs stop running if the session cluster's Job Manager is restarted without High Availability configured. However, because of a Flink bug, such stopped jobs get stuck in RECONCILING/STABLE state and cannot be restarted or deleted.

In such cases, the following is seen when using the kubectl get FlinkSessionJobs -n flink command:

```
kubectl get FlinkSessionJobs -n flink
NAME                          JOB STATUS    LIFECYCLE STATE
ssb-ssbdefault-testjobname    RECONCILING   STABLE
```

1. Remove the `FlinkDeployment` resource related to Cloudera SQL Stream Builder:

```
kubectl -n [***NAMESPACE***] delete flinkdeployment [***DEPL
OYMENT NAME***]
```

2. Open the Cloudera SQL Stream Builder configurations to edit:

```
kubectl -n [***NAMESPACE***] edit cm ssb-config
```

3. Add the highlighted line to the security-context.yaml entry:

```
allowPrivilegeEscalation: false
  capabilities:
    drop:
    - ALL
  runAsNonRoot: true
  seccompProfile:
    type: RuntimeDefault
  runAsUser: 9999
```

   It can take a couple of minutes for the changes to take effect after modifying the `ConfigMap`.
4. Execute the SQL job from Cloudera SQL Stream Builder again.

FLINK-33536: S3 filesystem sink and CSV format throws error

When using the Flink Table API CSV streaming sink with the S3 filesystem, the operation fails with IOException: Stream closed.

Cloudera SQL Stream Builder

CSADS-85: Data sampling might not start automatically after job reaches RUNNING

In some cases, data sampling does not start automatically when a job transitions to the RUNNING state.

Restart the sampler or restart the job manually.

Flink

DataStream conversion limitations

- Converting between Tables and POJO DataStreams is currently not supported in Cloudera Streaming Analytics Operator for Kubernetes.
- 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.

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 behavior 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

Learn what features are unsupported in this release of Cloudera Streaming Analytics Operator for Kubernetes.

Some Apache Flink and Cloudera SQL Stream Builder features exist in Cloudera Streaming Analytics Operator for Kubernetes, but are not supported by Cloudera. The following features are 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.

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

Cloudera SQL Stream Builder

- Virtual environments for Python are not supported
- Javascript UDFs are not supported

Deprecation notices in Cloudera Streaming Analytics Operator for Kubernetes

Certain features and functionalities have been removed or deprecated in Cloudera Streaming Analytics Operator for Kubernetes. 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 Operator for Kubernetes release. Marking an item as deprecated gives you time to plan for removal in a future Cloudera Streaming Analytics Operator for Kubernetes release.

Moving

Technology that Cloudera is moving from a future Cloudera Streaming Analytics Operator for Kubernetes 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 Operator for Kubernetes release and plan for the alternative Cloudera offering or subscription for the technology.

Removed

Technology that Cloudera has removed from Cloudera Streaming Analytics Operator for Kubernetes 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 Cloudera SQL Stream Builder

Certain features and functionality are deprecated or removed in Cloudera SQL Stream Builder. You must review these changes along with the information about the features in Cloudera SQL Stream Builder that will be removed or deprecated in a future release.

Removed

v1 REST API

The v1 REST API for Cloudera SQL Stream Builder is removed in Cloudera Streaming Analytics Operator for Kubernetes 1.4.0 and higher.

Cloudera recommends migrating to the v2 API.

For more information on the v2 API, see the Cloudera SQL Stream Builder REST API reference.

Support for JavaScript UDFs

User-Defined Functions (UDFs) written in JavaScript are removed in Cloudera Streaming Analytics Operator for Kubernetes 1.4.0.

Cloudera recommends using Python UDFs and migrating JavaScript UDFs to Python. Support in Python UDFs for table transformations and the webhook connector is available for Cloudera SQL Stream Builder and Flink.

Component support

You can review the Cloudera Streaming Analytics Operator for Kubernetes components and their versions shipped in this release of theCloudera Streaming Analytics Operator for Kubernetes.

Table 1: Cloudera Streaming Analytics Operator for Kubernetes component versions

Component	Version
Flink	1.20.1-csaop1.5.0-b275
Flink Operator	1.13-csaop1.5.0-b275
Cloudera SQL Stream Builder	1.20.1-csaop1.5.0-b275
PostgreSQL	18.1
OpenJDK	17

Supported Flink versions

Cloudera Streaming Analytics Operator for Kubernetes supports the following Flink versions:

Table 2: Supported Flink versions

Latest (default)	Other
1.20.1-csaop1.5.0-b275	None

The default version is the Cloudera-recommended current and latest supported Flink version. This version is used by default to deploy clusters if an explicit version is not provided in your `FlinkDeployment` resource.

Notice that the Flink versions are specific to Cloudera. Their version numbers consist of two parts: the first three digits specify the Apache Flink version, and the following specify the major and minor version of Cloudera Streaming Analytics Operator for Kubernetes. When deploying a cluster, you must use the Cloudera versions for Flink listed here. Specifying upstream versions is not supported.

Component support

Flink file system support

By default, Cloudera Streaming Analytics Operator for Kubernetes only supports local and locally mounted NFS/SAN file systems for Flink. Pluggable file systems (for example S3, HDFS, etc.) can be used by adding plugins to the Apache Flink Operator. For more information and a list of supported pluggable file systems, see *Using pluggable file systems*.

Cloudera SQL Stream Builder database support

You can use the following databases with Cloudera SQL Stream Builder:

- MySQL/MariaDB
- Oracle
- PostgreSQL

Cloudera SQL Stream Builder connector support

With Cloudera SQL Stream Builder, you will get the following connectors installed by default:

- Kafka
- JDBC
- CDC (MySQL, Oracle, Postgres, Db2, SqlServer)
- Amazon S3
- Azure Blob Storage
- Google Cloud Storage
- HDFS (when using the alternative image, see Installation overview)
- Hive (when using the alternative image, see Installation overview)
- Iceberg (when using the alternative image, see Installation overview)
- Kudu (when using the alternative image, see Installation overview)

You will also get the following formats installed with Cloudera SQL Stream Builder:

- JSON
- Avro
- ORC
- Parquet

Related Information

Using pluggable file systems

System requirements

Cloudera Streaming Analytics Operator for Kubernetes requires a Kubernetes cluster environment that meets the provided system requirements and prerequisites. You must ensure to meet these requirements to be able to install and use the Cloudera Streaming Analytics Operator for Kubernetes or its components.

- A Kubernetes 1.25 or later cluster:
  - OpenShift 4.12 or later
  - Any Cloud Native Computing Foundation (CNCF) certified Kubernetes distribution. For more information, see cncf.io.
- Administrative rights on the Kubernetes cluster.
- Access to kubectl or oc. These command line tools must be configured to connect to your running cluster.
- Access to helm.

- Log collection is enabled for the Kubernetes cluster.

  Cloudera requires that the logs of Cloudera Streaming Analytics Operator for Kubernetes components are stored long term for diagnostic and supportability purposes. Review Log collection.