Cloudera Streaming Analytics Operator 1.5.0

Cloudera Streaming Analytics Operator for Kubernetes Overview

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

https://docs.cloudera.com/

Legal Notice

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 is Cloudera Streaming Analytics Operator for Kubernetes?

Cloudera Streaming Analytics Operator for Kubernetes allows you to deploy and manage the Cloudera Streaming Analytics components of Cloudera, Flink, and Cloudera SQL Stream Builder, as container applications on Kubernetes.

Cloudera Streaming Analytics Operator for Kubernetes is based on the Apache Flink Kubernetes Operator (Flink Operator), and offers the same set of features as the Flink Operator supports. By extending the Kubernetes API, the Flink Operator acts as a control plane to manage the deployment lifecycle of Flink applications using the operator pattern. Cloudera Streaming Analytics Operator for Kubernetes also includes Cloudera SQL Stream Builder.

The operator pattern is a specialized controller used to manage a custom resource. As a customized control loop, controllers watch the shared state of the cluster through an API server to make changes in order to reach the desired state of the deployment defined by the custom resources. Building on default Kubernetes components, custom resources enable customization of the Kubernetes installation specialized for the application. This enables the Flink deployment process to be automated to start clusters, deploy stateful jobs, and ensure stability.

Note:

Due to the potential for confusion between the different meanings of "operator," references in the Cloudera Streaming Analytics Operator for Kubernetes documentation use the following terminology:

- Cloudera Streaming Analytics Operator for Kubernetes: refers to the Cloudera Streaming Analytics Operator for Kubernetes product
- Flink Operator: refers to the Apache Flink Kubernetes Operator component
- operator: refers to the function of the Flink to transform one or more DataStreams into a new DataStream.

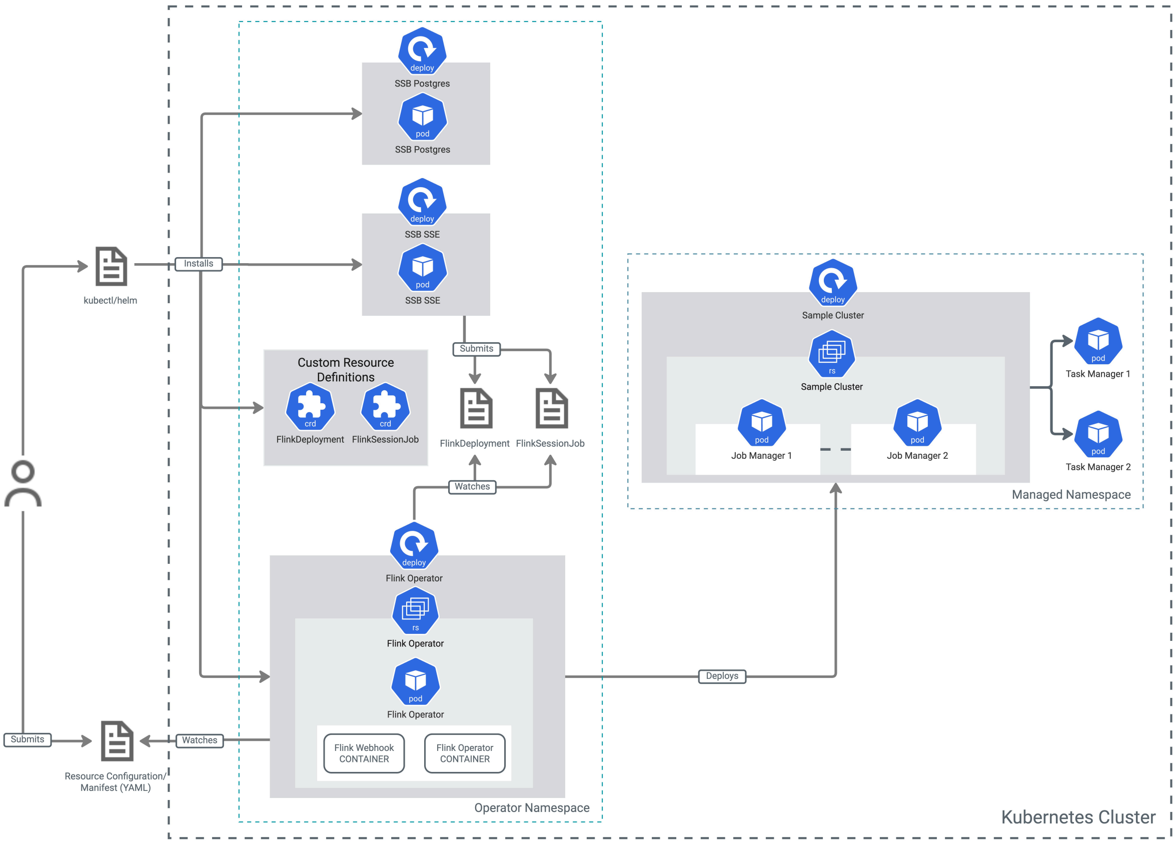
Related Information

DataStream Operators | Apache Flink

Deployment architecture

Cloudera Streaming Analytics Operator for Kubernetes can be installed on a Kubernetes cluster using Helm. The installation of Cloudera Streaming Analytics Operator for Kubernetes creates the deployment of the Apache Flink Kubernetes Operator (Flink Operator), Flink Webhook, and registers the Custom Resource Definitions (CRDs) on the Kubernetes cluster. As an extension, built on top of the Flink Operator, Cloudera Streaming Analytics Operator for Kubernetes also deploys the Cloudera SQL Stream Builder engine and its corresponding PostgreSQL database.

The Flink Operator is deployed in a designated namespace. After installation, the Flink Deployment CRD is registered, responsible for bringing up the Flink cluster. Flink deployments are controlled in one or more managed namespaces by the Flink Operator. When you deploy a new Flink deployment, the JobManager pod is created alongside with different Configmaps for the Flink Operator to function. Submitting a Flink deployment with a Flink job will deploy the required TaskManagers for the job to start. The following diagram shows the deployment architecture of the Flink Operator:



When installing the Cloudera Streaming Analytics Operator for Kubernetes with Helm, the Flink Operator Webhook is also installed as a custom admission plugin, which allows dynamic admission control. Similarly to connectors, you can use it to add plugins to the Flink operator that add custom rules triggered by certain actions.

There are two types of webhooks:

- mutating webhook: if you want to automatically configure some values on it or even force certain config values whenever a user creates a new `FlinkDeployment`, you can create a `FlinkResourceMutator`. Whenever a new `FlinkDeployment` is submitted, Kubernetes will call the webhook of the operator, and apply the custom mutator on the deployment.
- validating webhook: with this type, you cannot apply any changes to the deployment, but can automatically reject the creation of the deployment by implementing custom rules via the webhook.

The Flink Operator Webhook uses the TLS protocol to communicate by default, and automatically loads/reloads the keystore file when the file changes.

SSB integration

Cloudera Streaming Analytics Operator for Kubernetes comes with seamless Cloudera SQL Stream Builder integration. Cloudera SQL Stream Builder is built on top of the Flink Operator, offering an interactive user interface for creating streaming SQL jobs.

Cloudera SQL Stream Builder is a comprehensive interactive user interface for creating stateful stream processing jobs using SQL. Using SQL, you can simply and easily declare expressions that filter, aggregate, route, and otherwise mutate streams of data. Cloudera SQL Stream Builder offers a job management interface that you can use to compose and run SQL on streams, as well as to create durable data APIs for the results.

The Helm chart contains the Cloudera SQL Stream Builder subchart, which has two deployments in the Cloudera Streaming Analytics Operator for Kubernetes: `sse`, which provides the engine and User Interface (UI), and `postgres`, which provides the default database for Cloudera SQL Stream Builder to function. When submitting a SQL job using the UI (Streaming SQL Console), the parsed SQL is serialized, compressed and encrypted into an environment and a Flink job is deployed. This means that, under the hood, Cloudera SQL Stream Builder creates the same Flink Deployment as you would for a generic Flink job, with the only difference that a special Flink job is created for the SQL Runner. The SQL Runner decrypts and decompresses the parsed SQL, sets up the environment inside the Flink job, and executes it. Cloudera SQL Stream Builder by default deploys the jobs in Session Mode. You can use the installed connectors in your SQL jobs as a source or sink with the supported data formats.

Related Information

Architecture | Apache Flink Kubernetes Operator

Licensing

Cloudera Streaming Analytics Operator for Kubernetes requires a valid license to function. Licenses are made available for you, together with your Cloudera credentials, as part of your license and subscription agreement with Cloudera.

Licenses are registered during Cloudera Streaming Analytics Operator for Kubernetes installation. They are stored in a Kubernetes secret. Licenses can be updated at any time.

Licenses are valid for a set period of time. Once the license expires, the cluster resources you deployed will continue to run, but reconciliation of resources will be blocked. For example: failed pods will not be restarted and scaling your clusters will not be possible. In general, the control mechanisms in place that keep resources healthy will be blocked. This will result in deployed resources breaking down over time.

Cloudera Streaming Analytics Operator for Kubernetes publishes various log entries and Kubernetes events related to your licenses.

For example, if your license expires or becomes invalid for any reason, relevant logs and events published will notify you that there are issues with your license.

These logs and events are published for the Flink Operator deployment. You can check these logs and events with the following commands:

```
kubectl events deployments/csa-operator --namespace [***NAMESPACE***]
```

```
kubectl logs deployment/csa-operator --namespace [***NAMESPACE***]
```

Related Information

Updating Cloudera license