Cloudera Streaming Analytics Operator 1.0.0

CSA Operator Installation

Date published: 2024-06-15
Date modified: 2024-06-15

https://docs.cloudera.com/

Legal Notice

Contents

Installation overview

Get started with installing CSA Operator. Learn about the concept of installing CSA Operator, the installation artifacts, and where these artifacts are hosted.

CSA Operator is installed using a Helm chart, which installs the Apache Kubernetes Flink Operator (Flink Operator), SQL Stream Builder (SSB), and other components.

Installation artifacts and artifact locations

CSA Operator ships with various installation artifacts, hosted at two locations: the Cloudera Docker registry and the Cloudera Archive.

Both the Cloudera Docker registry and the Cloudera Archive require your Cloudera credentials (username and password) to access. Credentials are provided to you as part of your license and subscription agreement. You can access both the registry and the archive using the same credentials.

Cloudera Docker registry – container.repository.cloudera.com

The Cloudera Docker registry hosts the Helm chart as well as all Docker images used for the installation.

Table 1: CSA Operator artifacts on the Cloudera Docker registry

Artifact	Location	Description
Flink Kubernetes Operator Docker image	container.repository.cloudera.com/cloudera/flink-kubernetes-operator:1.8-csaop1.0.0-b317	Docker image used for deploying the various operator components shipped with the CSA Operator.
Flink Docker image	container.repository.cloudera.com/cloudera/flink:1.18.1-csaop1.0.0-b317	Docker image used for deploying Apache Flink and its related components.
SQL Runner Docker image	container.repository.cloudera.com/cloudera/ssb-sql-runner:1.18.1-csaop1.0.0-b317	Docker image used for deploying Flink application when SQL query is executed in SSB.
SQL Stream Engine Docker image	container.repository.cloudera.com/cloudera/ssb-sse:1.18.1-csaop1.0.0-b317	Docker image used for deploying SSB and the SSB UI.

Cloudera Archive – archive.cloudera.com/p/csa-operator/1.0.0/

The Cloudera Archive hosts various installation artifacts including the Helm chart, diagnostic tools, and the Maven artifacts.

All artifacts hosted in the Cloudera Archive are supplemental resources, accessing them is not required to complete the installation. The following table collects the CSA Operator directories located in the Cloudera Archive, with an overview of what artifacts they contain and how you can use them:

Table 2: CSA Operator artifacts on the Cloudera Archive

Archive Directory	Description
archive.cloudera.com/p/csa-operator/1.0.0/**charts**/	The charts directory contains the Helm chart. This is the same chart that is available in the Cloudera Docker registry. Cloudera recommends that whenever possible you install with the chart hosted in the registry. The chart in the archive is provided in case you cannot access the registry or want to download the chart using a browser.

Archive Directory	Description
archive.cloudera.com/p/csa-operator/1.0.0/**maven-repository**/	The Maven artifacts can be used to develop your own applications or tools for use with CSA Operator.
archive.cloudera.com/p/csa-operator/1.0.0/**tools**/	The tools directory contains command line tools that you use to collect diagnostic information and to troubleshoot cluster issues.

Installing CSA Operator in an environment with internet access

Complete these steps to install CSA Operator if your Kubernetes cluster has internet access. Installing the CSA Operator enables you to deploy and manage Flink and SQL Stream Builder (SSB) in Kubernetes.

About this task

CSA Operator is installed in your Kubernetes cluster with the provided Helm chart through the helm install command. When you install the chart, Helm installs the Custom Resource Descriptors (CRD) included in CSA Operator, and deploys the Apache Kubernetes Flink Operator (Flink Operator), SSB engine (in Technical Preview), and a PostgreSQL database for SSB.

Installing CSA Operator does not create or deploy a Flink cluster. The Flink cluster is created after the installation by deploying the Flink Deployment resource in the Kubernetes cluster with kubectl or oc, or when you execute a SQL job in Streaming SQL Console.

Note: Cloudera recommends that you install CSA Operator once per Kubernetes cluster.

By default the Flink Operator (deployed with installation) watches and manages all the Flink clusters that are deployed in the single same namespace as the Flink Operator. However, you can also configure it to watch and manage multiple namespaces. This allows you to manage multiple Flink clusters deployed in different namespaces, using a single CSA Operator installation.

Important: SSB can only be deployed in one namespace. In case you require multiple instances of SSB, you must install SSB in every namespace. For more information about deploying Flink and SSB in multiple namespaces, see the Namespace management documentation.

Before you begin

- Ensure that your Kubernetes environment meets requirements listed in System requirements.
- Your Kubernetes cluster requires internet connectivity to complete these steps, as it must be able to reach the Cloudera Docker registry.
- Ensure that you have access to a valid Cloudera license.
- Ensure that you have access to your Cloudera credentials (username and password, that are provided together with the Cloudera license) required to access the Cloudera Docker registry (and, if needed, the Cloudera Archive), where installation artifacts are hosted.
- Review the Helm chart reference before installation.

  The Helm chart accepts various configuration properties that you can set during installation. You can customize your installation using these properties.
- If you want to use the Webhook of Flink Operator, ensure that you have cert-manager installed on your Kubernetes cluster, which you can install using the following command:

```
kubectl create -f https://github.com/jetstack/cert-manager/releases/down
load/v1.8.2/cert-manager.yaml
```

```
kubectl wait -n cert-manager --for=condition=Available deployment --all
```

- The webhook functionality is enabled by default. You can disable it using the following command, and skip the cert-manager installation:

  ```
  --set flink-kubernetes-operator.webhook.create=false
  ```

Procedure

1. Create a namespace in your Kubernetes cluster where you will install and use the CSA Operator.

   ```
   kubectl create namespace [***NAMESPACE***]
   ```

   This is the namespace where you install Flink and SSB. Use this namespace you create in all installation steps that follow.
2. Create a Kubernetes secret to contain your Cloudera credentials.

   ```
   kubectl create secret docker-registry [***SECRET NAME***] \
       --docker-server container.repository.cloudera.com \
       --docker-username [***USERNAME***] \
       --docker-password [***PASSWORD***] \
       --namespace [***NAMESPACE***]
   ```

   Ensure that the placeholders are replaced with your specific information:

   a) Provide a desired name for *[***SECRET NAME***]*.
   b) Replace *[***USERNAME***]* and *[***PASSWORD***]* with your Cloudera credentials.
   c) Provide the same name for *[***NAMESPACE***]* that you created in the previous step.
3. Log in to Cloudera Docker registry with helm.

   ```
   helm registry login container.repository.cloudera.com
   ```

   Enter your Cloudera credentials when prompted.
4. Install CSA Operator with helm install.

   ```
   helm install csa-operator --namespace [***NAMESPACE***] \
       --set 'flink-kubernetes-operator.imagePullSecrets[0].name=[***SECRET
    NAME***]' \
       --set 'ssb.sse.image.imagePullSecrets[0].name=[***SECRET NAME***]' \
       --set 'ssb.sqlRunner.image.imagePullSecrets[0].name=[***SECRET
    NAME***]' \
       --set-file flink-kubernetes-operator.clouderaLicense.fileCon
   tent=[***PATH TO LICENSE FILE***] \
   oci://container.repository.cloudera.com/cloudera-helm/csa-operator/csa-op
   erator --version 1.0.0-b317
   ```

   Important: When you install CSA Operator, SSB will also be installed by default. SSB is in Technical Preview in this version of the CSA Operator. In case you want to skip installing SSB, add --set ssb.enabled=false to the helm install command.

   Ensure that the placeholders are replaced with your specific information:

   a) Provide the same name for *[***NAMESPACE***]* that you created in Step 1.
   b) Provide the same name for *[***SECRET NAME***]* that you created in the previous step. imagePullSecrets specifies what secret is used to pull images from the Cloudera registry. Setting this property is mandatory, otherwise Helm will be unable to pull the necessary images from the Cloudera Docker registry.
   c) Replace *[***PATH TO LICENSE FILE***]* with the full (absolute) path to your Cloudera license file. clouderaLicense.fileContent is used to register your license. When this property is set, a secret is generated that

contains the license you specify. Setting this property is mandatory. The CSA Operator will not function without a valid license.

d) You can use --set to set various other properties of the Helm chart. This enables you to customize your installation. For example, by default the Flink Operator has access to watch all namespaces. However, you can configure a list of specific namespaces to watch using watchNamespaces.. For example, in case you created multiple namespaces, you can configure the Flink Operator to only watch specific ones with --set flink-kubernetes-operator.watchNamespaces={*[***NAMESPACE1***], [***NAMESPACE2***]*}. For more information about deploying Flink and SSB in multiple namespaces, see the Namespace management documentation.

5. Check that the Flink Operator, and the SSB engine with its database are running.

```
kubectl get pods -n [***NAMESPACE***]
```

```
NAME                                READY   STATUS      RESTARTS    AGE
flink-kubernetes-operator           1/2     Running      0          7s
ssb-postgresql                      1/1     Running      0          7s
ssb-sse                             1/1     Running      0          7s
```

What to do next

After successfully installing the CSA Operator, you can start using Flink and SSB (in Technical Preview) on Kubernetes. The Getting Started with Flink and Getting Started with SSB guides can help you with the basic operations.

Installing CSA Operator in an air-gapped environment

Complete these steps to install CSA Operator if your Kubernetes cluster does not have internet access, or if you want to install from a self-hosted registry. Installing the CSA Operator enables you to deploy and manage Flink and SQL Stream Builder (SSB) in Kubernetes.

About this task

CSA Operator is installed in your Kubernetes cluster with the provided Helm chart through the helm install command. When you install the chart, Helm installs the Custom Resource Descriptors (CRD) included in CSA Operator, and deploys the Apache Kubernetes Flink Operator (Flink Operator), SSB engine (in Technical Preview), and a PostgreSQL database for SSB.

Installing CSA Operator does not create or deploy a Flink cluster. The Flink cluster is created after the installation by deploying the Flink Deployment resource in the Kubernetes cluster with kubectl or oc, or when you execute a SQL job in Streaming SQL Console.

Note: Cloudera recommends that you install CSA Operator once per Kubernetes cluster.

By default the Flink Operator (deployed with installation) watches and manages all the Flink clusters that are deployed in the single same namespace as the Flink Operator. However, you can also configure it to watch and manage multiple namespaces. This allows you to manage multiple Flink clusters deployed in different namespaces, using a single CSA Operator installation.

Important: SSB can only be deployed in one namespace. In case you require multiple instances of SSB, you must install SSB in every namespace. For more information about deploying Flink and SSB in multiple namespaces, see the Namespace management documentation.

Before you begin

- Ensure that your Kubernetes environment meets requirements listed in System requirements.
- A self-hosted Docker registry is required. Your registry must be accessible by your Kubernetes cluster.

- While the Kubernetes cluster does not need internet access in an air-gapped environment, the preparation steps to create the local (offline) repository, from which you can install CSA Operator, require that you can download and move the artifacts hosted on the Cloudera Docker registry and Cloudera Archive.
- Access to docker or equivalent utility that you can use to pull and push images is required. The Cloudera-recommended way is using docker. Replace commands where necessary, if you use a different utility.
- Ensure that you have access to your Cloudera credentials (username and password). Credentials are required to access the Cloudera Docker registry (and, if needed, the Cloudera Archive) where installation artifacts are hosted.
- Ensure that you have access to a valid Cloudera license.
- Review the Helm chart reference before installation.

  The Helm chart accepts various configuration properties that you can set during installation. Using these properties you can customize your installation.
- If you want to use the Webhook of Flink Operator, ensure that you have cert-manager installed on your Kubernetes cluster, which you can install using the following command:

```
kubectl create -f https://github.com/jetstack/cert-manager/releases/down
load/v1.8.2/cert-manager.yaml
kubectl wait -n cert-manager --for=condition=Available deployment --all
```

  - The webhook functionality is enabled by default. You can disable it using the following command, and skip the cert-manager installation:

```
--set flink-kubernetes-operator.webhook.create=false
```

Procedure

1. Copy the following installation artifacts to your self-hosted registry.

Table 3: CSA Operator artifacts on the Cloudera Docker registry

Artifact	Location	Description
Flink Kubernetes Operator Docker image	container.repository.cloudera.com/cloudera/flink-kubernetes-operator:1.8-csaop1.0.0-b317	Docker image used for deploying the various operator components shipped with the CSA Operator.
Flink Docker image	container.repository.cloudera.com/cloudera/flink:1.18.1-csaop1.0.0-b317	Docker image used for deploying Apache Flink and its related components.
SQL Runner Docker image	container.repository.cloudera.com/cloudera/ssb-sql-runner:1.18.1-csaop1.0.0-b317	Docker image used for deploying Flink application when SQL query is executed in SSB.
SQL Stream Engine Docker image	container.repository.cloudera.com/cloudera/ssb-sse:1.18.1-csaop1.0.0-b317	Docker image used for deploying SSB and the SSB UI.

   This step involves pulling the artifacts from the Cloudera Docker registry, retagging them, and then pushing them to your self-hosted registry. The exact steps you need to carry it out depend on your environment and how you set up your registry. The following substeps demonstrate a basic workflow using docker and helm.

   a) Log in to the Cloudera Docker registry with both docker and helm.

      Provide your Cloudera credentials when prompted.

```
docker login container.repository.cloudera.com
```

```
helm registry login container.repository.cloudera.com
```

   b) Pull the Docker images from the Cloudera Docker registry.

```
docker pull \
```

```
  container.repository.cloudera.com/cloudera/[***IMAGE
 NAME***]:[***VERSION***]
```

c) Pull the CSA Operator Helm chart.

```
helm pull \
  oci://container.repository.cloudera.com/cloudera-helm/csa-operator/csa
-operator \
  --version 1.0.0-b317
```

d) Retag the Docker images you pulled so that they contain the address of your registry.

```
docker tag \
[***ORIGINAL IMAGE TAG***] \
[***REGISTRY HOSTNAME***]:[***PORT***]/cloudera/[***IMAGE NAME***]:
[***VERSION***]
```

e) Push the images and chart to your self-hosted registry.

```
docker push \
[***REGISTRY HOSTNAME***]:[***PORT***]/cloudera/[***IMAGE
 NAME***]:[***VERSION***]
```

```
helm push \
csa-operator-1.0.0-b317.tgz \
oci://[***REGISTRY HOSTNAME***]:[***PORT***]/cloudera-helm/csa-operator/
```

2. Create a namespace in your Kubernetes cluster where you will install and use the CSA Operator.

```
kubectl create namespace [***NAMESPACE***]
```

This is the namespace where you install Flink and SSB. Use this namespace you create in all installation steps that follow.

3. Create a Kubernetes secret to credentials for your self-hosted registry.

```
kubectl create secret docker-registry [***SECRET NAME***] \
    --docker-server [***REGISTRY HOSTNAME***]:[***PORT***] \
    --docker-username [***USERNAME***] \
    --docker-password [***PASSWORD***] \
    --namespace [***NAMESPACE***]
```

Ensure that the placeholders are replaced with your specific information:

a) Provide a desired name for *[***SECRET NAME***]*.

b) Replace *[***REGISTRY HOSTNAME***]*:*[***PORT***]* with your self-hosted registry hostname and port.

c) Replace *[***USERNAME***]* and *[***PASSWORD***]* with your Cloudera credentials.

d) Provide the same name for *[***NAMESPACE***]* that you created in the previous step.

4. Log in to your self-hosted registry with helm.

```
helm registry login [***REGISTRY HOSTNAME***]:[***PORT***]
```

Enter your credentials when prompted.

5. Install CSA Operator with helm install.

```
helm install csa-operator \
    --namespace [***NAMESPACE***] \
    --set 'flink-kubernetes-operator.image.repository=[***REGISTRY
 HOSTNAME***]:[***PORT***]/cloudera/[***IMAGE NAME***]'\
    --set 'ssb.sqlRunner.image.repository=[***REGISTRY
 HOSTNAME***]:[***PORT***]/cloudera/[***IMAGE NAME***]'\
```

```
    --set 'ssb.sse.image.repository=[***REGISTRY
 HOSTNAME***]:[***PORT***]/cloudera/[***IMAGE NAME***]'\
    --set 'flink-kubernetes-operator.imagePullSecrets[0].name=[***SECRET
 NAME***]' \
    --set 'ssb.sse.image.imagePullSecrets[0].name=[***SECRET NAME***]' \
    --set 'ssb.sqlRunner.image.imagePullSecrets[0].name=[***SECRET
 NAME***]' \
    --set-file flink-kubernetes-operator.clouderaLicense.fileCon
tent=[***PATH TO LICENSE FILE***] \
oci://[***REGISTRY HOSTNAME***]:[***PORT***]/cloudera-helm/csa-operator/
csa-operator --version 1.0.0-b317
```

Important: When you install CSA Operator, SSB will also be installed by default. SSB is in Technical Preview in this version of the CSA Operator. In case you want to skip installing SSB, add --set ssb.enabled=false to the helm install command.

Ensure that the placeholders are replaced with your specific information:

a) Provide the same name for *[***NAMESPACE***]* that you created in Step 1.

b) Replace *[***REGISTRY HOSTNAME***]*:*[***PORT***]* with your self-hosted registry hostname and port.

c) Provide the same name for *[***SECRET NAME***]* that you created in the previous step. imagePullSecrets specifies what secret is used to pull images from the Cloudera registry. Setting this property is mandatory, otherwise, Helm cannot pull the necessary images from the Cloudera Docker registry.

d) Replace *[***PATH TO LICENSE FILE***]* with the full (absolute) path to your Cloudera license file. clouderaLicense.fileContent is used to register your license. When this property is set, a secret is generated that contains the license you specify. Setting this property is mandatory. The CSA Operator will not function without a valid license.

e) You can use --set to set various other properties of the Helm chart. This enables you to customize your installation. For example, by default the Flink Operator has access to watch all namespaces. However, you can configure a list of specific namespaces to watch using watchNamespaces.. For example, in case you created multiple namespaces, you can configure the Flink Operator to only watch specific ones with --set flink-kubernetes-operator.watchNamespaces={*[***NAMESPACE1***], [***NAMESPACE2***]*}. For more information about deploying Flink and SSB in multiple namespaces, see the Namespace management documentation.

6. Check that the Flink Operator, and the SSB engine with its database are running.

```
kubectl get pods -n [***NAMESPACE***]
```

```
NAME                                READY   STATUS      RESTARTS    AGE
flink-kubernetes-operator           1/2     Running      0          7s
ssb-postgresql                      1/1     Running      0          7s
ssb-sse                             1/1     Running      0          7s
```

What to do next

After successfully installing the CSA Operator, you can start using Flink and SSB (in Technical Preview) on Kubernetes. The Getting Started with Flink and Getting Started with SSB guides can help you with the basic operations.