

Installation

Date published: 2019-12-17

Date modified: 2019-12-17



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

Installing CSD and parcel.....	4
Adding Flink as a Service.....	6
Setting up your HDFS Home directory.....	8
Setting the Java executable for the Flink client.....	9

Installing CSD and parcel

For installing Cloudera Streaming Analytics (CSA), you need to add the downloaded Flink Custom Service Descriptor (CSD) file to the default CSD directory, and add the parcel to your cluster using Cloudera Manager.

Before you begin

- Download the CSD and parcel file.
- Install CDP Private Cloud Base. For more information about installing CDP Private Cloud Base and Cloudera Manager, see the CDP Private Cloud Base documentation.
- Check that the following components are installed:

Required components	Optional components
YARN	Kafka
HDFS	HBase
Zookeeper	Schema Registry
	Streams Messaging Manager

Procedure

1. Place the CSD file in the `/opt/cloudera/csd/` folder (default CSD directory). This way Cloudera Manager will automatically detect the CSD file.
2. Change the ownership of the CSD file.

```
> chown cloudera-scm:cloudera-scm FLINK-1.9.1-csa1.1.0.0-cdh7.0.3.0-79-1753674.jar
```

3. Restart Cloudera Manager and CMS services for the changes to take effect.

```
> systemctl restart cloudera-scm-server
```

4. Log into Cloudera Manager.
5. Select Parcels on the Home > Hosts tab in the main navigation bar.
6. Click the Parcel Repositories & Network Settings tab.
7. Check that Cloudera Manager successfully auto-detected the Remote Parcel Repository URL for CSA.



Note: Make sure that the Remote Parcel Repository URL uses HTTPS link.

8. Enter your download credentials to HTTP authentication username override for Cloudera Repositories and HTTP authentication password override for Cloudera

CLUSTERA Manager Parcels

Search

Cluster Hosts Diagnostics Audit Chart Replication Administration

Parcel Repository & Network Settings

Cloudera Manager checks the connection to the configured repository.

> 9/9 URL(s) - The repository was successfully accessed

Remote Parcel Repository URLs

Uninstall

Enable Automatic Authentication for Cloudera Repositories ☒

HTTP authentication username override for Cloudera Repositories

Uninstall

HTTP authentication password override for Cloudera Repositories

Uninstall

Proxy Server

Proxy Port

CDEP Deployment 01:12

Parcel Running

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
You will be redirected to the Parcels page.
11. Search for Flink and click Download to download the parcel to the local repository.
12. After the download is completed, click Distribute to distribute the parcel to all clusters.
13. After the parcel is distributed, click Activate to activate the parcel.
14. Click OK when confirmation is required.

For more details, follow the standard [procedure](#) from the GUI or the API.

Results

You have added the Flink CSD file and parcel to your cluster. You can test your installation using the following command on any node in your cluster:

```
> flink --version
...
Version: 1.9.1-csa1.1.0.0, Commit ID: 8e2dd0c
```

What to do next

Add Flink as a service in Cloudera Manager.

Related Information

[Adding Flink as a Service](#)

[Installing CDP Private Cloud Base](#)

Adding Flink as a Service

You need to use the Add Service wizard in Cloudera Manager to have Flink service on your cluster. When assigning roles, you must install Flink, HDFS and YARN Gateway roles on the same node from where the Flink jobs are submitted.

Before you begin

You have placed the CSD file to the /opt/cloudera/csd folder and installed the Flink parcel on Cloudera Manager.

Procedure

1. Open Cloudera Manager.
2. On the Home screen, select the drop-down menu to the right of your cluster.
3. Select Add Service.
4. From the list, select Flink as the type of service, then click Continue.
The Add Service wizard will launch.

5. Assign roles to the History Server and Gateway, then click

The screenshot shows the Cloudera Manager interface. On the left is a dark blue sidebar with the Cloudera Manager logo at the top. At the bottom of the sidebar, there are two menu items: 'Parcels' with a gift icon and 'Recent Commands' with a document icon. The main area on the right is titled 'Add Flink Service'. Below the title is a vertical progress bar with five steps: 1. Select Dependencies (marked with a blue checkmark), 2. Assign Roles (highlighted with a blue circle and bold text), 3. Review Changes, 4. Command Details, and 5. Summary. Below the progress bar, the text 'CDEP Deployment from 2020-Jan-17 01:40' is visible.

CLUDERA Manager

Add Flink Service

- ✓ Select Dependencies
- 2 Assign Roles**
- 3 Review Changes
- 4 Command Details
- 5 Summary

CDEP Deployment from 2020-Jan-17
01:40

Parcels

Recent Commands

**Note:**

Install Flink, HDFS, and YARN Gateway roles on the same machine that will be used to submit Flink jobs. The Flink HistoryServer role also depends on having HDFS client configurations on the same machine. The HDFS client configurations can either be provided by an HDFS daemon role implicitly or can be deployed by an HDFS Gateway role explicitly.

- Review the changes needed for your service.

**Note:**

You can leave this page blank as the settings are configured automatically. You can later change the security settings at [Flink > Configuration > Security](#).

- Click Continue and wait until the first run of the Flink service is completed.
- Click Continue and then Finish.

Results

You have added Flink as a service in Cloudera Manager.

Setting up your HDFS Home directory

You need an HDFS Home directory to store temporary logs and data of your application to run a Flink job. You must set up the HDFS home directory for your user to avoid error when using Flink.

About this task

To run a Flink job, your HDFS Home Directory has to exist. If it does not exist, you receive an error message similar to:

```
Permission denied: user=$USER_NAME, access=WRITE, inode="/user":
```

Procedure

Create HDFS Home directory. Ask your HDFS administrator to perform the following (or obtain [HDFS administrator role](#)).

Options**Command****Kerberos enabled**

```
> kinit hdfs
> hdfs dfs -mkdir /user/$USER_NAME
> hdfs dfs -chown $USER_NAME:$USER_NAME /user/$USER_NAME
```

Kerberos disabled

```
> HADOOP_USER_NAME=hdfs hdfs dfs -mkdir /user/$USER_NAME
> HADOOP_USER_NAME=hdfs hdfs dfs -chown $USER_NAME:$USER_NAME /user/$USER_NAME
```

In case of an enterprise environment, you can use Hue to set up the Home directory by enabling automatic synchronization for users. For more information see, the Cloudera Runtime [documentation](#).

Setting the Java executable for the Flink client

You must set the `Java_home` environment through the command line for the Flink clients manually to avoid error when using Flink. The configuration in Cloudera Manager only applies to services, and not to clients.

About this task

Cloudera Manager offers a configuration for the `JAVA_HOME` environment variable under `Hosts > All Hosts > Configuration`. However, this only applies to services (for example YARN NodeManager or Flink HistoryServer) and does not propagate to clients such as the JVM created locally by the Flink executable. JVM uses the Bigtop utility under `/usr/bin/bigtop-detect-javahome` to automatically detect the `JAVA_HOME`.

Procedure

Set `JAVA_HOME` to a fixed value:

```
> cat /etc/default/bigtop-utils
export JAVA_HOME=/usr/java/default
```



Note:

Cloudera strongly recommends setting this to the same value as set in Cloudera Manager. It is also recommended to set it uniformly on all the nodes to avoid unnecessary confusion. This is a known issue in the [Cloudera Community](#).