Cloudera Streaming Analytics

Installation

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

https://docs.cloudera.com/

Legal Notice

Contents

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.

Note: 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
	Kudu
	Atlas

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-1.10.0-csa1.2.0.0-cdh7.1.1.0-565
-3525501.jar
```

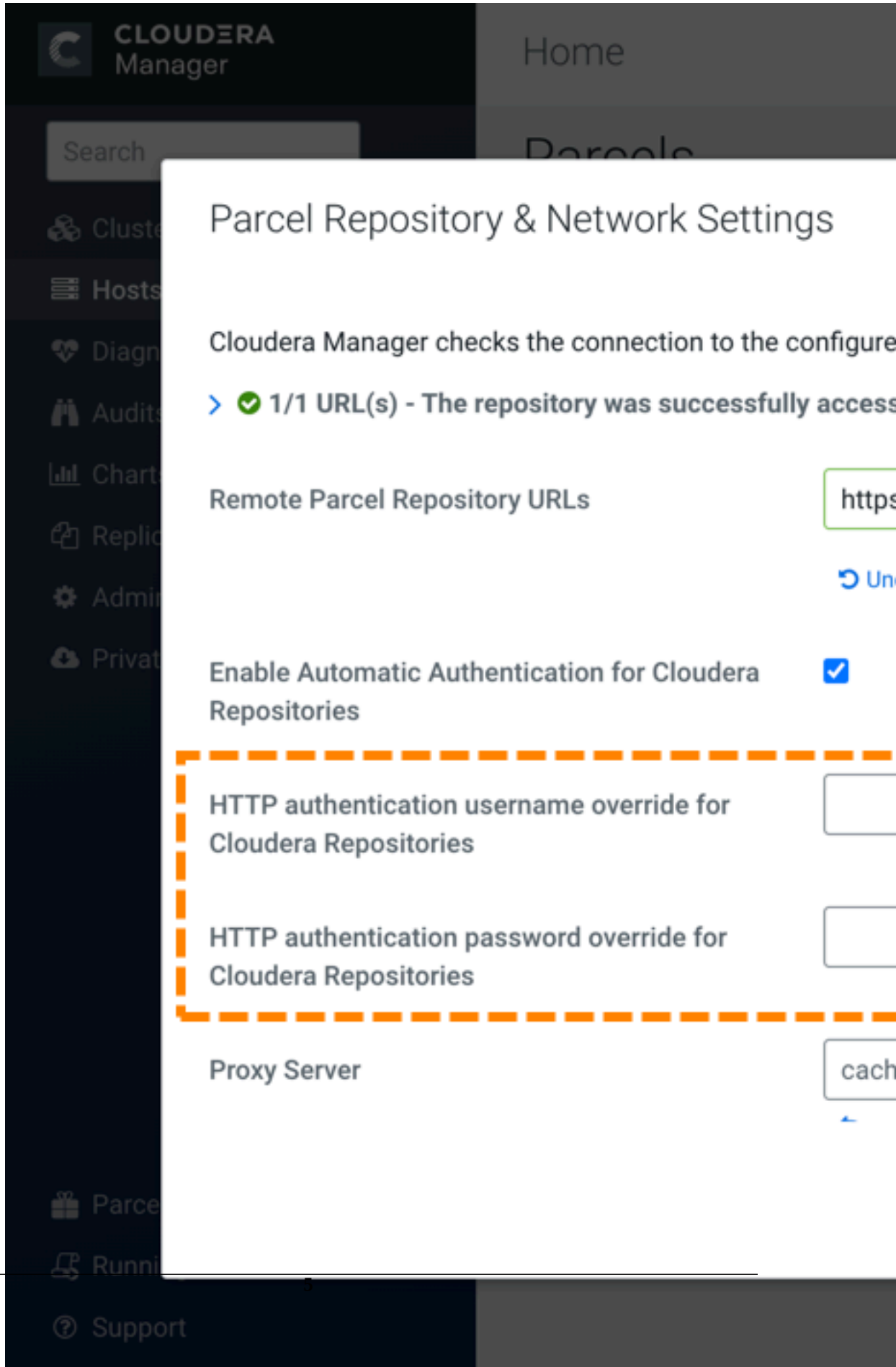
3. Restart Cloudera Manager and CMS services for the changes to take effect.

```
> 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. In case you need to install a different version of the parcel, you can change the URL as needed.

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



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: 1.10.0-csa1.2.0.0, Commit ID: 04dddd1
```

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. Choose HBase and Hive as Optional dependency if needed for the source and sink solution, then click Continue.
6. Assign roles to the History Server and Gateway, then click Continue.

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.

7. Review the changes needed for your service.

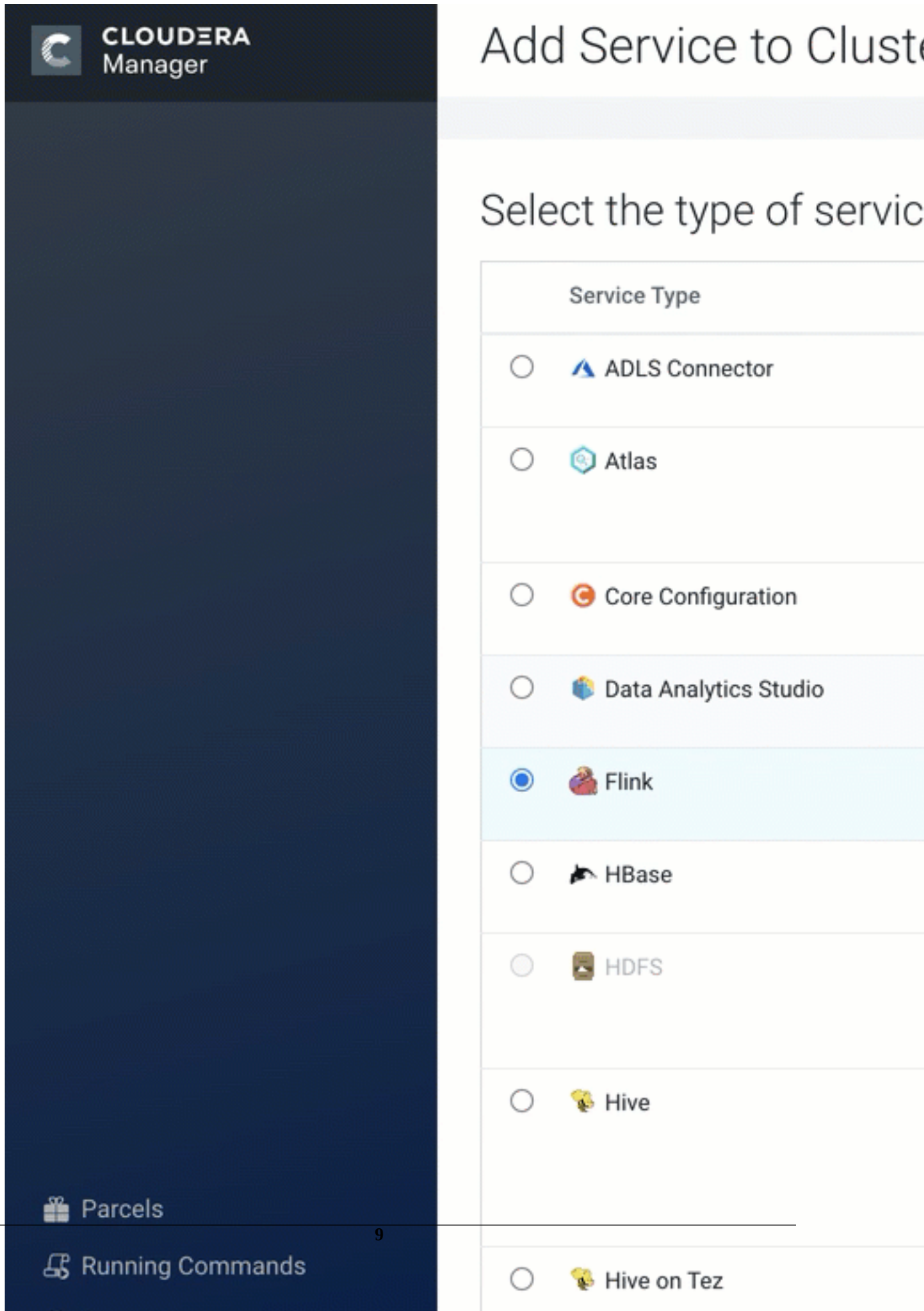
   

   Note:

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

8. Click Continue and wait until the first run of the Flink service is completed.

9. Click Continue and then Finish.


CLOUDERA
Manager
Add Service to Cluste
Select the type of servic
Service Type
ADLS Connector
Atlas
Core Configuration
Data Analytics Studio
Flink
HBase
HDFS
Hive
Hive on Tez
Parcels
Running Commands

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.