Cloudera Manager

Managing Clusters

Date published: 2020-11-30
Date modified: 2026-02-27

https://docs.cloudera.com/

Legal Notice

Contents

Managing Cloudera Runtime Services

Cloudera Manager service configuration features let you manage the deployment and configuration of Cloudera Runtime and managed services.

Using Cloudera Manager, you can gracefully start, stop and restart services or roles. Further, you can modify the configuration properties for services or for individual role instances. . You can also generate client configuration files, enabling you to easily distribute them to the users of a service.

The topics in this chapter describe how to configure and use the services on your cluster. Some services have unique configuration requirements or provide unique features. See the documentation for an individual service for more information.

Adding a Service

After using the Cloudera Management Console to set up a cluster, you can add additional services.

About this task

When you add a service to a cluster in Cloudera on cloud, you must note the following points:

- The default configuration of new services may not match with the Cloudera on cloud environment, and the new service may not start up successfully. To find the correct configuration parameters for Cloudera on cloud, check the service configuration parameters in the Cloudera-provided cluster templates.
- Using the Cloudera Manager Add Service feature to manually add services can be error-prone and non-reproducible. Cloudera suggests using a cluster template to configure services, so that the configuration parameters for each service are captured, making it simple to re-create the cluster with all services.
- You can use Cloudera-provided cluster templates, or create custom cluster templates by starting with a Cloudera-provided cluster template as a base.

For more information on cluster templates, see Cluster templates overview.

When adding a service in Cloudera on cloud, you use the Host Templates created during set up of the cluster to add role groups for each of the roles in a service. You cannot add a role directly to a host. After adding the appropriate role group to the Host Template, you can scale up the cluster by applying the Host Template to the host. You can then re-apply the Host Template to all hosts, or you can scale up the cluster by applying a Host Template to the new hosts.

Each role should be assigned to a Host Template that is configured for master, worker, or gateway roles. Templates for master roles usually assign the role to a single host. Templates for worker or gateway hosts can assign roles to multiple hosts.

Minimum Required Role: Limited Cluster Administrator (also provided by Full Administrator and Cluster Administrator)

Procedure

To add a service:

1. On the HomeStatus tab, click  to the right of the cluster name and select Add a Service. A list of service types display. You can add one type of service at a time.
2. Select a service and click Continue. If you are missing required binaries, a pop-up displays asking if you want to continue with adding the service.
3. Select the services on which the new service should depend. All services must depend on the same ZooKeeper service. Click Continue.
   The Assign Templates page displays.

4. A list of roles displays on the page. For each role, select a Host template for the role. To see the current allocation of roles and role groups, click the View By Host Template button.
5. Click Continue
6. Review and modify configuration settings, such as data directory paths and heap sizes and click Continue. The service is started.

   Note: If you are adding the Ranger service, passwords for the Ranger Admin, Usersync, Tagsync, and KMS Keyadmin users must be a minimum of 8 characters long, with at least one alphabetic and one numeric character. The following characters are not valid: " ' \ ` ´.
7. Click Continue then click Finish. You are returned to the home page.
8. Verify the new service is started properly by checking the health status for the new service. If the Health Status is Good, then the service started properly.

Starting a Cloudera Runtime service on all hosts

Starting and stopping Cloudera Runtime services.

About this task

Minimum Required Role: Operator (also provided by Configurator, Cluster Administrator, Limited Cluster Administrator , and Full Administrator)

It is important to start and stop services that have dependencies in the correct order. For example, because MapReduce and YARN have a dependency on HDFS, you must start HDFS before starting MapReduce or YARN. The Cloudera Management Service and Hue are the only two services on which no other services depend; although you can start and stop them at anytime, their preferred order is shown in the following procedures. The Cloudera Manager cluster actions start and stop services in the correct order. To start or stop all services in a cluster, follow the instructions in Starting, Stopping, Refreshing, and Restarting a Cluster.

Before you begin

The order in which to start services is:

1. Cloudera Management Service
2. ZooKeeper
3. HDFS
4. Solr
5. HBase
6. Key-Value Store Indexer
7. MapReduce or YARN
8. Hive
9. Impala
10. Oozie
11. Sqoop
12. Hue

Procedure

1. In the left menu, click Clusters and select a service.
2. Click  to the right of the service name and select Start.
3. Click Start in the next screen to confirm.
   When you see a Finished status, the service has started.

Results

Note: If you are unable to start the HDFS service, it's possible that one of the roles instances, such as a DataNode, was running on a host that is no longer connected to the Cloudera Manager Server host, perhaps because of a hardware or network failure. If this is the case, the Cloudera Manager Server will be unable to connect to the Cloudera Manager Agent on that disconnected host to start the role instance, which will prevent the HDFS service from starting. To work around this, you can stop all services, abort the pending command to start the role instance on the disconnected host, and then restart all services again without that role instance.

Related Information

Aborting a Pending Command

Stopping a Cloudera Runtime Service on All Hosts

About this task

Minimum Required Role: Operator (also provided by Configurator, Cluster Administrator, Limited Cluster Administrator , and Full Administrator)

Before you begin

The order in which to stop services is:

1. Hue
2. Sqoop
3. Oozie
4. Impala
5. Hive
6. MapReduce or YARN
7. Key-Value Store Indexer
8. HBase
9. Flume
10. Solr
11. HDFS
12. ZooKeeper
13. Cloudera Management Service

Procedure

1. In the left menu, click Clusters and select a service.
2. Click  to the right of the service name and select Stop.
3. Click Stop in the next screen to confirm.
   When you see a Finished status, the service has stopped.

Restarting a Cloudera Runtime Service

About this task

Minimum Required Role: Operator (also provided by Configurator, Cluster Administrator, Limited Cluster Administrator , and Full Administrator)

Procedure

1. In the left menu, click Clusters and select a service.
2. Click  to the right of the service name and select Restart.
3. Click Start on the next screen to confirm.

Results

When you see a Finished status, the service has restarted.

What to do next

To restart all services, restart the cluster.

Rolling Restart

Rolling restart allows you to conditionally restart the role instances of the following services to update software or use a new configuration.

Minimum Required Role: Operator (also provided by Configurator, Cluster Administrator, Limited Cluster Administrator , and Full Administrator)

The following services support Rolling Restart operations:

- Atlas
- HBase
- HDFS
- Hive-on-Tez
- Hive Metastore
- Hue
- Kafka
- Knox
- Kudu – see Orchestrating a rolling restart with no downtime.
- MapReduce
- OMID
- Oozie
- Phoenix
- Ranger KMS
- Schema Registry
- Spark
- Spark 3
- Streams Replication Manager
- YARN
- ZooKeeper

If the service is not running, rolling restart is not available for that service. You can specify a rolling restart of each service individually.

If you have HDFS High Availability enabled, you can also perform a cluster-level rolling restart. At the cluster level, the rolling restart of worker hosts is performed on a host-by-host basis, rather than per service, to avoid all roles for a service potentially being unavailable at the same time. During a cluster restart, to avoid having your NameNode (and thus the cluster) be unavailable during the restart, Cloudera Manager forces a failover to the standby NameNode.

Job Tracker and Resource Manager High availability are not required for a cluster-level rolling restart. However, if you have JobTracker or ResourceManager high availability enabled, Cloudera Manager will force a failover to the standby JobTracker or ResourceManager.

Performing a Service or Role Rolling Restart

You can initiate a rolling restart from either the Status page for one of the eligible services, or from the service's Instances page, where you can select individual roles to be restarted.

1. Go to the service you want to restart.
2. Do one of the following:
   - service - Select ActionsRolling Restart.
   - role -
     a. Click the Instances tab.
     b. Select the roles to restart.
     c. Select Actions for SelectedRolling Restart.
3. In the pop-up dialog box, select the options you want:
   - Restart only roles whose configurations are stale
   - Restart only roles that are running outdated software versions
   - Which role types to restart
4. If you select an HDFS, HBase, MapReduce, or YARN service, you can have their worker roles restarted in batches. You can configure:
   - How many roles should be included in a batch - Cloudera Manager restarts the worker roles rack-by-rack in alphabetical order, and within each rack, hosts are restarted in alphabetical order. If you are using the default replication factor of 3, Hadoop tries to keep the replicas on at least 2 different racks. So if you have multiple racks, you can use a higher batch size than the default 1. But you should be aware that using too high batch size also means that fewer worker roles are active at any time during the upgrade, so it can cause temporary performance degradation. If you are using a single rack only, you should only restart one worker node at a time to ensure data availability during upgrade.
   - How long should Cloudera Manager wait before starting the next batch.

- The number of batch failures that will cause the entire rolling restart to fail (this is an advanced feature). For example if you have a very large cluster you can use this option to allow failures because if you know that your cluster will be functional even if some worker roles are down.

Note:

- HDFS - If you do not have HDFS high availability configured, a warning appears reminding you that the service will become unavailable during the restart while the NameNode is restarted. Services that depend on that HDFS service will also be disrupted. Cloudera recommends that you restart the DataNodes one at a time—one host per batch, which is the default.
- HBase
  - Administration operations such as any of the following should not be performed during the rolling restart, to avoid leaving the cluster in an inconsistent state:
    - Split
    - Create, disable, enable, or drop table
    - Metadata changes
    - Create, clone, or restore a snapshot. Snapshots rely on the RegionServers being up; otherwise the snapshot will fail.
  - To increase the speed of a rolling restart of the HBase service, set the Region Mover Threads property to a higher value. This increases the number of regions that can be moved in parallel, but places additional strain on the HMaster.
  - Another option to increase the speed of a rolling restart of the HBase service is to set the Skip Region Reload During Rolling Restart property to true. This setting can cause regions to be moved around multiple times, which can degrade HBase client performance.
- MapReduce - If you restart the JobTracker, all current jobs will fail.
- OMID - Cloudera recommends not to disable the High Availability (HA) mode for the OMID Transactional Status Oracle (TSO) server. In case you want to disable the HA for the OMID TSO server, ensure that only one instance is running before disabling it.

  OMID constantly checks the service status with Zookeeper because OMID depends on Zookeeper service while in HA mode. This increases the network traffic and might impact the service performance.
- YARN - If you restart ResourceManager and ResourceManager HA is enabled, current jobs continue running: they do not restart or fail.
- ZooKeeper and Flume - For both ZooKeeper and Flume, the option to restart roles in batches is not available. They are always restarted one by one.

5. Click Confirm to start the rolling restart.

Performing a Cluster-Level Rolling Restart

You can perform a cluster-level rolling restart on demand from the Cloudera Admin Console. A cluster-level rolling restart is also performed as the last step in a rolling upgrade when the cluster is configured with HDFS high availability enabled.

1. If you have not already done so, enable high availability. See HDFS High Availability for instructions. You do not need to enable automatic failover for rolling restart to work, though you can enable it if you want. Automatic failover does not affect the rolling restart operation.
2. For the cluster you want to restart select ActionsRolling Restart.
3. In the pop-up dialog box, select the services you want to restart. Please review the caveats in the preceding section for the services you elect to have restarted. The services that do not support rolling restart will simply be restarted, and will be unavailable during their restart.
4. If you select an HDFS, HBase, or MapReduce service, you can have their worker roles restarted in batches. You can configure:

   - How many roles should be included in a batch - Cloudera Manager restarts the worker roles rack-by-rack in alphabetical order, and within each rack, hosts are restarted in alphabetical order. If you are using the default

replication factor of 3, Hadoop tries to keep the replicas on at least 2 different racks. So if you have multiple racks, you can use a higher batch size than the default 1. But you should be aware that using too high batch size also means that fewer worker roles are active at any time during the upgrade, so it can cause temporary performance degradation. If you are using a single rack only, you should only restart one worker node at a time to ensure data availability during upgrade.

- How long should Cloudera Manager wait before starting the next batch.
- The number of batch failures that will cause the entire rolling restart to fail (this is an advanced feature). For example if you have a very large cluster you can use this option to allow failures because if you know that your cluster will be functional even if some worker roles are down.

Note: All Hue roles (namely the load balancer, Kerberos ticket renewer, and Hue server) are of the non-worker type. You must select one of the following options under the Roles to include section for ensuring that the Hue service is restarted in the rolling restart mode, depending on your requierments:

- Non-Workers Only, or
- All Roles

5. Click Restart to start the rolling restart. While the restart is in progress, the Command Details page shows the steps for stopping and restarting the services.

Aborting a Pending Command

Minimum Required Role: Operator (also provided by Configurator, Cluster Administrator, Limited Cluster Administrator , and Full Administrator)

Commands will time out if they are unable to complete after a period of time.

If necessary, you can abort a pending command. For example, this may become necessary because of a hardware or network failure where a host running a role instance becomes disconnected from the Cloudera Manager Server host. In this case, the Cloudera Manager Server will be unable to connect to the Cloudera Manager Agent on that disconnected host to start or stop the role instance which will prevent the corresponding service from starting or stopping. To work around this, you can abort the command to start or stop the role instance on the disconnected host, and then you can start or stop the service again.

To abort any pending command:

You can click the Recent Commands indicator (), which shows the number of commands that are currently running in your cluster (if any). This indicator is positioned above the Support link at the bottom of the left menu. Unlike the Commands tab for a role or service, this indicator includes all commands running for all services or roles in the cluster. In the **Running Commands** window, click Abort to abort the pending command.

To abort a pending command for a service or role:

1. In the left menu, click Clusters and select the service where the role instance you want to stop is located. For example, click ClustersHDFS Service if you want to abort a pending command for a DataNode.
2. Click the Instances tab.
3. In the list of instances, click the link for role instance where the command is running (for example, the instance that is located on the disconnected host).
4. Go to the Commands tab.
5. Find the command in the list of Running Commands and click Abort Command to abort the running command.

Related Information

Viewing Running and Recent Commands

Deleting Services

You can delete a service from the **Status** tab.

About this task

Minimum Required Role: Limited Cluster Administrator (also provided by Full Administrator and Cluster Administrator)

Procedure

1. Stop the service.
2. On the HomeStatus tab, click  to the right of the service name and select Delete.
3. Click Delete to confirm the deletion.

   Deleting a service does not clean up the associated client configurations that have been deployed in the cluster or the user data stored in the cluster. For a given "alternatives path" (for example /etc/hadoop/conf) if there exist both "live" client configurations (ones that would be pushed out with deploy client configurations for active services) and ones that have been "orphaned" client configurations (the service they correspond to has been deleted), the orphaned ones will be removed from the alternatives database.

   You must check the dependencies for a service to be deleted in the service matrix. For more information, see Service Dependencies in Cloudera Manager.

   Note that to trigger cleanup of client configurations associated with a deleted service you must create a service to replace it. To remove user data, see the topic *Remove Cloudera Manager and User Data*.