

Cloudera Manager

Managing Clusters

Date published: 2024-12-10

Date modified: 2024-12-10

CLOUDERA

<https://docs.cloudera.com/>

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

Pausing a Cluster in AWS.....	4
Shutting Down and Starting Up the Cluster.....	4
Managing Cloudera Runtime Services.....	5
Adding a Service.....	5
Prerequisites for installing Atlas.....	7
Installing Atlas using Add Service.....	7
Installing Ranger using Add Service.....	10
Comparing Configurations for a Service Between Clusters.....	13
Starting a Cloudera Runtime Service on All Hosts.....	13
Stopping a Cloudera Runtime Service on All Hosts.....	14
Restarting a Cloudera Runtime Service.....	15
Rolling Restart.....	15
Aborting a Pending Command.....	18
Deleting Services.....	18
Renaming a Service.....	19
Configuring Maximum File Descriptors.....	19
Extending Cloudera Manager.....	20
Add-on Services.....	20
Configuring Services to Use LZO Compression.....	22

Pausing a Cluster in AWS

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

If all data for a cluster is stored on EBS volumes, you can pause the cluster and stop your AWS EC2 instances during periods when the cluster will not be used. The cluster will not be available while paused and can't be used to ingest or process data, but you won't be billed by Amazon for the stopped EC2 instances. Provisioned EBS storage volumes will continue to accrue charges.



Important: Pausing a cluster requires using EBS volumes for all storage, both on management and worker nodes. Data stored on ephemeral disks will be lost after EC2 instances are stopped.

Shutting Down and Starting Up the Cluster

To pause an AWS cluster, follow the shutdown procedure. To restart the cluster after a pause, follow the startup procedure.



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

In the shutdown and startup procedures below, some steps are performed in the AWS console and some are performed in Cloudera Manager:

- For AWS actions, use one of the following interfaces:
 - AWS console
 - AWS CLI
 - AWS API
- For cluster actions, use one of the following interfaces:
 - The Cloudera Manager web UI
 - The Cloudera API start and stop commands



Shutdown procedure

To pause the cluster, complete the following steps:

1. Navigate to the Cloudera Manager web UI.
2. Stop the cluster.
 - a. On the HomeStatus tab, click  to the right of the cluster name and select Stop.
 - b. Click Stop in the confirmation screen. The Command Details window shows the progress of stopping services.
When All services successfully stopped appears, the task is complete and you can close the Command Details window.
3. Stop the Cloudera Management Service.
 - a. On the HomeStatus tab, click  to the right of the service name and select Stop.
 - b. Click Stop in the next screen to confirm. When you see a Finished status, the service has stopped.
4. In AWS, stop all cluster EC2 instances, including the Cloudera Manager host.

Startup procedure

To restart the cluster after a pause, the steps are reversed:

1. In AWS, start all cluster EC2 instances.
2. Navigate to the Cloudera Manager UI.
3. Start the Cloudera Management Service.
 - a. On the HomeStatus tab, click  to the right of the service name and select Start.
 - b. Click Start that appears in the next screen to confirm. When you see a Finished status, the service has started.
4. Start the cluster.
 - a. On the HomeStatus tab, click  to the right of the cluster name and select Start.
 - b. Click Start that appears in the next screen to confirm. The Command Details window shows the progress of starting services.

When All services successfully started appears, the task is complete and you can close the Command Details window.

Considerations after Restart

Since the cluster was completely stopped before stopping the EC2 instances, the cluster should be healthy upon restart and ready for use. You should be aware of the following about the restarted cluster:

- After starting the EC2 instances, Cloudera Manager and its agents will be running but the cluster will be stopped. There will be gaps in Cloudera Manager's time-based metrics and charts.
- EC2 instances retain their internal IP address and hostname for their lifetime, so no reconfiguration of CDH or Runtime is required after restart. The public IP and DNS hostnames, however, will be different. Elastic IPs can be configured to remain associated with a stopped instance at additional cost, but it isn't necessary to maintain proper cluster operation.

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 add new services and roles if needed, gracefully start, stop and restart services or roles, and decommission and delete roles or services if necessary. Further, you can modify the configuration properties for services or for individual role instances. You can also view past configuration changes and roll back to a previous revision. 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 initial installation, you can use the **Add a Service** wizard to add and configure new service instances. For example, you may want to add a service such as Oozie that you did not select in the wizard during the initial installation.

About this task

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

**Note:**

The binaries for the following services are not packaged in Cloudera Runtime and must be installed individually before being adding the service:

- Accumulo
- Kafka
- Key Trustee KMS

If you do not add the binaries before adding the service, the service will fail to start.




Important: Additional Steps are required if the cluster is running a Solr service and you want to add either the Apache Atlas or Apache Ranger services:

1. Follow the steps below to [add a service](#) and add an additional Solr Service.
2. When the Add a Service wizard prompts you to fill in configuration properties, set the ZooKeeper Znode property to solr-infra.
3. Complete adding the Solr service as described below.
4. Run the Add a Service wizard again to add the Ranger or Atlas service.
5. Cloudera Manager will prompt you for the Solr service to use. Select the additional Solr service you just added.

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 Roles page displays
4. Customize the assignment of role instances to hosts. The wizard evaluates the hardware configurations of the hosts to determine the best hosts for each role. The wizard assigns all worker roles to the same set of hosts to which the HDFS DataNode role is assigned. You can reassign role instances.

Click a field below a role to display a dialog box containing a list of hosts. If you click a field containing multiple hosts, you can also select All Hosts to assign the role to all hosts, or Custom to display the hosts dialog box.

The following shortcuts for specifying hostname patterns are supported:

- Range of hostnames (without the domain portion)

Range Definition	Matching Hosts
10.1.1.[1-4]	10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4
host[1-3].company.com	host1.company.com, host2.company.com, host3.company.com
host[07-10].company.com	host07.company.com, host08.company.com, host09.company.com, host10.company.com

- IP addresses
- Rack name

Click the View By Host button for an overview of the role assignment by hostname ranges.

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.

Prerequisites for installing Atlas

Using Cloudera Manager, you can install Apache Atlas and there are some prerequisites that you must note before you proceed further.

Services that are part of prerequisites to install Atlas include:

- HBase: Acts as storage for Atlas data.
- Solr: Acts as storages for indexes of Atlas data.
- Kafka: Acts as a messaging bus for Atlas hooks.
- Ranger: Works with Atlas to provide tag-based authorization.

Also verify if your Cloudera Manager instance has PAM Authentication enabled. For more information, see [Configure Atlas PAM Authentication](#). Atlas installation can be performed if your Cloudera Manager cluster has PAM Authentication enabled and it works with the users enabled on host.

The following authentication options are available for the installation:

- File-based authentication.
- LDAP or Active Directory (AD): LDAP and AD can be configured after Atlas is installed.
- PAM-based Authentication based installation. For more information, see [Configure Atlas PAM Authentication](#).
- Knox SSO. This needs Knox Service to be present.

Depending on your usage scenario the following authentication options can be chosen:

- Production deployment:
 - PAM Authentication
 - LDAP and AD
 - Knox SSO
- Demo and testing purposes:
 - File-based authentication

Related Information

[Installing Atlas using Add Service](#)

Installing Atlas using Add Service

You can install Atlas service in your Cloudera Manager instance by using the Add Service method.

Procedure

1. From Cloudera Manager > select Add Service menu option.
2. On the Add Service page, Select Atlas and click Continue. You will be navigated to a new listing page that prompts you to follow the next set of instructions.
3. Choose Select Dependencies.

Make sure that there are no additional dependencies required on your Cloudera Manager cluster.

4. Click Continue.

5. Select Assign Roles by selecting the host where Atlas is to be added. You must also add the Gateway (optional) but recommended.
6. Select Continue.

You must assign roles and also make sure you do not assign additional roles that can impact the overall performance.

Add Atlas Service to Cluster 1

Select Dependencies

2 **Assign Roles**

3 Migrate Navigator Data

4 Enable Migration Mode

5 Review Changes

6 Command Details

7 Summary

Assign Roles

You can customize the role assignments for your new service here, but note that if assignments are made incorrectly, such as assigning too many roles to a single host, performance will suffer.

You can also view the role assignments by host. [View By Host](#)

Atlas Server × 1 New

Gateway × 3 New

vs-6082102-2.vs-6082102.root.hwx.site

vs-6082102-[1-3].vs-6082102.root.hwx.site ▾

Cancel

← Back

Continue →

7. Review the changes and ensure that the following fields (depending on the installation type that you have chosen) are marked as checked. For more information, review the Atlas installation prerequisite section.

- (Optional) Enable Kerberos Authentication - Atlas (Service-Wide)
- (Optional) Enable File Authentication - Atlas Service Default Group

You must add the Admin Password as applicable.

Add Atlas Service to Cluster 1

The 'Review Changes' dialog displays the following configuration options:

- Enable Kerberos Authentication:** Atlas (Service-Wide)
- Admin Password:** Atlas Server Default Group
- Enable File Authentication:** Atlas Server Default Group
- Excluded Wire Encryption Protocols:** Atlas Server Default Group (TLSv1, TLSv1.1)
- Initial Solr Shards for Atlas Collections:** Atlas Server Default Group (1)
- Initial Solr Replication Factor for Collections:** Atlas Server Default Group (1)
- Atlas Max Heapsize:** Atlas Server Default Group (2 GB)
- Kafka Message Retention Time:** Atlas Server Default Group (31 days)
- Server HTTP port:** Atlas Server Default Group (31000)
- Server HTTPS Port:** Atlas Server Default Group (31443)
- Enable TLS/SSL for Atlas Server:** Atlas Server Default Group
- Atlas Server TLS/SSL Server JKS Keystore File Location:** Atlas Server Default Group
- Atlas Server TLS/SSL Server JKS Keystore File Password:** Atlas Server Default Group

Buttons: Cancel, Back, Continue

8. Review the Command Details and make sure that the installation proceeds on the expected lines.

9. Click Continue.

The 'Start' dialog shows the following progress:

- Status: **Running** (Context: Atlas, Jun 10, 10:02:13 AM)
- Completed 0 of 1 step(s).
- Starting 1 roles on service: 0/1 start commands completed.
- Execute command Start this Atlas Server on role Atlas Server (vs-6082102-2): Atlas Server (vs-6082102-2) (Jun 10, 10:02:13 AM)
- Start a role: Role is starting. Atlas Server (vs-6082102-2) (Jun 10, 10:02:13 AM)

Buttons: Abort, Close

When using a fresh Cloudera Manager cluster, collections are created directly.

Summarise the installation process to confirm that Atlas service is installed and configured on your Cloudera Manager cluster.

10. Once the installation is complete, to bring up Atlas on your Cloudera Manager cluster, restart the Atlas service.

11. Once Atlas is installed, you must restart stale configurations.

Related Information

[Prerequisites for installing Atlas](#)

[Deploying Atlas service](#)

Installing Ranger using Add Service

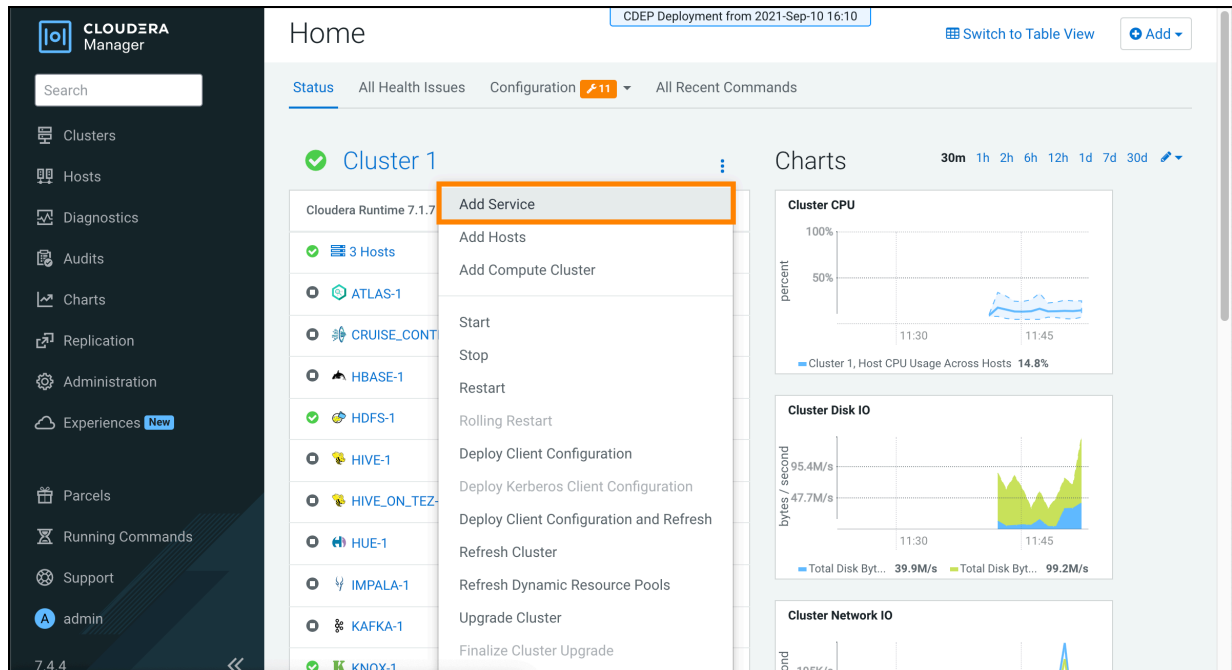
You can install Ranger service in an existing cluster using the Add Service action.

Before you begin

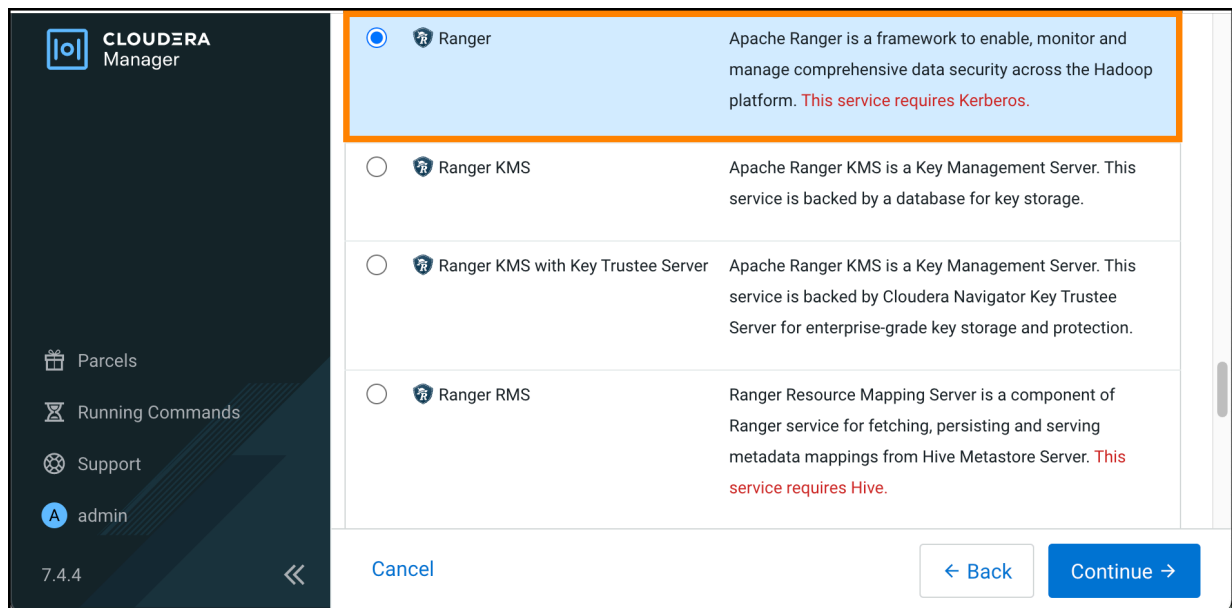
- Ranger requires a [supported database](#). This should be a separate database than the one used for Ranger KMS. For more information, see [Configuring a database for Ranger or Ranger KMS](#).
- [TLS/SSL should be enabled](#).

Procedure

1. On the cluster home page, click More Options (ellipsis icon), then click Add Service.



2. Select Ranger, then click Continue.



3. On Assign Roles, click Continue.

The screenshot shows the 'Assign Roles' step in Cloudera Manager. The left sidebar indicates the progress: 1. Select Dependencies (checked), 2. Assign Roles (active), 3. Review Changes, 4. Command Details, and 5. Summary. The main content area is titled 'Assign Roles' and includes instructions: 'You can customize the role assignments for your new service here, but note that if assignments are made incorrectly, such as assigning too many roles to a single host, performance will suffer.' Below this, it says 'You can also view the role assignments by host.' with a 'View By Host' button. Three role assignment cards are visible: 'Ranger Admin x 1 New', 'Usersync x 1 New', and 'Ranger Tagsync x 1 New'. At the bottom, there are 'Cancel', 'Back', and 'Continue' buttons.

4. On Setup Database, enter the Ranger host name, database name, user name, and password, then click Test Connection. After the database connection is successful, click Continue.

Field Name	Description
Database Host Name	The host name of the Ranger database. If a non-default port is assigned, use host:port notation.
Database Name	The name of the Ranger database. Default value is ranger1
(Database) User Name	The Ranger database user name. Default value is rangeradmin
(Database) Password	The Ranger database user password. Default value is rangeradmin

The screenshot shows the 'Setup Database' step in Cloudera Manager. The left sidebar indicates the progress: 1. Select Dependencies (checked), 2. Assign Roles (checked), 3. Setup Database (active), 4. Review Changes, 5. Command Details, and 6. Summary. The main content area is titled 'Setup Database' and includes instructions: 'Configure and test database connections. If using custom databases, create the databases first according to the [Installing and Configuring an External Database](#) section of the [Installation Guide](#).' Below this, the 'Ranger' configuration section has the following fields: 'Type' (MySQL), 'Use JDBC URL Override' (No), 'Database Hostname' (ranger.db.hostname), 'Database Name' (ranger1), 'Username' (rangeradmin), and 'Password' (rangeradmin). There is a 'Show Password' checkbox (checked) and a 'Test Connection' button. A 'Notes' section contains the following information:

- The value in the **Database Hostname** field must match the value you used for the hostname when creating the database.
- If the database is not running on its default port, specify the port number using **host:port** in the **Database Hostname** field.
- It is highly recommended that each database is on the same host as the corresponding role instance.
- If a value in the **JDBC URL** field is provided, it will be used when establishing a connection to the database. This customized connection URL will override **Database Hostname**, **Type**, and **Database Name**. Only some services currently support this.
- [Learn more](#)

At the bottom, there are 'Cancel', 'Back', and 'Continue' buttons.

- On Review Changes, enter a master key password for Ranger Admin, Usersync, Tagsync, and KMS KeyAdmin, then click Continue.

Field Name	Description
Ranger Admin User Initial Password rangeradmin_user_password	Password for the Ranger administrator. Default value is admin123
Ranger Usersync User Initial Password rangerusersync_user_password	Password for the Ranger administrator. Default value is admin123
Ranger Tagsync User Initial Password rangertagsync_user_password	Password for the Ranger administrator. Default value is admin123
Ranger KMS Keyadmin User Initial Password keyadmin_user_password	Password for the Ranger administrator. Default value is admin123

- On Command Details, select run options, confirm success, then click Continue.

The screenshot shows the Cloudera Manager interface for the 'Add Ranger Service to Cluster 1' command. The left sidebar shows the navigation menu with 'Command Details' selected. The main content area displays the 'Command Details' for the 'First Run Command'. The status is 'Finished' and the context is 'Ranger'. The command was executed on Sep 11, 1:12:27 AM and took 12.95s. The command successfully executed the 'Setup Ranger Admin Component on service Ranger'. The 'Completed 1 of 1 step(s)' section shows a list of steps, including 'Run a set of services for the first time' and 'Execute 3 steps in sequence'.

- On Summary, click Finish.

The screenshot shows the Cloudera Manager interface for the 'Add Ranger Service to Cluster 1' command. The left sidebar shows the navigation menu with 'Summary' selected. The main content area displays the 'Summary' for the command. A green success message states: 'Your new service is installed and configured on your cluster.' A blue note box contains the following text: 'Note: You may still have to start your new service. It is recommended that you restart any dependency services with outdated configurations before doing so. You can perform these actions on the main page by clicking Finish below.'

- The Ranger service appears in the Cloudera Manager cluster components list. If Ranger was not started by the installation wizard, you can start the service by clicking **Actions** **Start** in the Ranger service.
- Restart all services with stale configurations.

10. In Cloudera Manager click the Ranger service, then select **Actions Create Ranger Plugin Audit Directory** . The Ranger service is now ready to use and you should be able to validate Ranger policy enforcement.

Comparing Configurations for a Service Between Clusters

You can compare the configuration settings for a particular service between two different clusters in a Cloudera Manager deployment.

About this task

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

Procedure

1. On the HomeStatus tab, click the name of the service you want to compare, or click the Clusters menu and select the name of the service.
2. Click the Configuration tab.
3. Click the drop-down menu above the Filters pane, and select from one of the options that begins Diff with...:
 - *service on cluster* - For example, HBASE-1 on Cluster 1. This is the default display setting. All properties are displayed for the selected instance of the service.
 - *service on all clusters* - For example, HBase on all clusters. All properties are displayed for all instances of the service.
 - *Diff with service on cluster* - For example, Diff with HBase on Cluster 2. Properties are displayed only if the values for the instance of the service whose page you are on differ from the values for the instance selected in the drop-down menu.
 - *Diff with service on all clusters* - For example, Diff with HBase on all clusters. Properties are displayed if the values for the instance of the service whose page you are on differ from the values for one or more other instances in the Cloudera Manager deployment.

The service's properties will be displayed showing the values for each property for the selected clusters. The filters on the left side can be used to limit the properties displayed.

You can also view property configuration values that differ between clusters across a deployment by selecting **Non-uniform Values** on the Configuration tab of the Cloudera Manager HomeStatus page.

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)

Before you begin

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

- Key Trustee Server
- 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 Manager 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 requirements:

- 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*.

Renaming a Service

A service is given a name upon installation, and that name is used as an identifier internally. However, Cloudera Manager allows you to provide a display name for a service, and that name will appear in the Cloudera Manager Admin Console instead of the original (internal) name.

About this task


Minimum Required Role: [Full Administrator](#). This feature is not available when using Cloudera Manager to manage Data Hub clusters.



Note:

The original service name will still be used internally, and may appear or be required in certain circumstances, such as in log messages or in the API.

Procedure

1. On the HomeStatus tab, click  to the right of the service name and select Rename.
2. Type the new name.
3. Click Rename.

The rename action is recorded as an Audit event.

When looking at Audit or Event search results for the renamed service, it is possible that these search results might contain either only the original (internal) name, or both the display name and the original name.

Configuring Maximum File Descriptors

You can set the maximum file descriptor parameter for all daemon roles. When not specified, the role uses whatever value it inherits from supervisor. When specified, configures soft and hard limits to the configured value.

About this task

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

Procedure

1. Go to a service.
2. Click the **Configuration** tab.
3. In the Search box, type `rlimit_fds`.
4. Set the Maximum Process File Descriptors property for one or more roles.
5. Enter a Reason for change, and then click Save Changes to commit the changes.
6. Restart the affected role instances.



Important:

After changing the maximum process file descriptor limit as specified in the above steps, you might find that service roles are still limited by the number of file descriptors. Raising the maximum process file descriptors above the Linux kernel file descriptor limit will have no effect. Check the Linux kernel file descriptor limit on every host in the cluster and raise that if necessary.

You can find the Linux kernel file descriptor limit by running the following command on the Linux command line:

```
sudo cat /proc/sys/fs/nr_open
```

Extending Cloudera Manager

In addition to the set of software packages and services managed by Cloudera Manager, you can also define and add new types of services using [custom service descriptors](#). When you deploy a custom service descriptor, the implementation is delivered in a Cloudera Manager [parcel](#) or other software package. For information on the extension mechanisms provided by Cloudera Manager for creating custom service descriptors and parcels, see [Cloudera Manager Extensions](#).

Add-on Services

Minimum Required Role: [Full Administrator](#). This feature is not available when using Cloudera Manager to manage Data Hub clusters.

Cloudera Manager supports adding new types of services (referred to as an *add-on service*) to Cloudera Manager, allowing such services to leverage Cloudera Manager distribution, configuration, monitoring, resource management, and life-cycle management features. An add-on service can be provided by Cloudera or an independent software vendor (ISV). If you have multiple clusters managed by Cloudera Manager, an add-on service can be deployed on any of the clusters.



Note: If the add-on service is already installed and running on hosts that are not currently being managed by Cloudera Manager, you must first add the hosts to a cluster that's under management. See [Adding a Host to a Cluster](#) for details.

Custom Service Descriptor Files

Integrating an add-on service requires a Custom Service Descriptor (CSD) file. A CSD file contains all the configuration needed to describe and manage a new service. A CSD is provided in the form of a JAR file.

Depending on the service, the CSD and associated software may be provided by Cloudera or by an ISV. The integration process assumes that the add-on service software (parcel or package) has been installed and is present on the cluster. The recommended method is for the ISV to provide the software as a parcel, but the actual mechanism for

installing the software is up to the ISV. The instructions in [Installing an Add-on Service](#) on page 21 assume that you have obtained the CSD file from the Cloudera repository or from an ISV. It also assumes you have obtained the service software, ideally as a parcel, and have or will install it on your cluster either prior to installing the CSD or as part of the CSD installation process.

Configuring the Location of Custom Service Descriptor Files

The default location for CSD files is `/opt/cloudera/csd`. You can change the location in the Cloudera Manager Admin Console as follows:

1. Select AdministrationSettings.
2. Click the Custom Service Descriptors category.
3. Edit the Local Descriptor Repository Path property.
4. Enter a Reason for change, and then click Save Changes to commit the changes.
5. Restart Cloudera Manager Server:

RHEL 7 compatible, SLES, Ubuntu:

```
sudo systemctl restart cloudera-scm-server
```

RHEL 6 compatible:

```
sudo service cloudera-scm-server restart
```


Installing an Add-on Service

An ISV may provide its software in the form of a parcel, or they may have a different way of installing their software. If their software is not available as a parcel, then you must install their software before adding the CSD file. Follow the instructions from the ISV for installing the software. If the ISV has provided their software as a parcel, they may also have included the location of their parcel repository in the CSD they have provided. In that case, install the CSD first and then install the parcel.

Installing the Custom Service Descriptor File

1. Acquire the CSD file from Cloudera or an ISV.
2. Log on to the Cloudera Manager Server host, and place the CSD file under the [location configured](#) for CSD files.
3. Set the file ownership to `cloudera-scm:cloudera-scm` with permission 644.
4. Restart the Cloudera Manager Server:

```
service cloudera-scm-server restart
```

5. Log into the Cloudera Manager Admin Console and restart the Cloudera Management Service.
 - a. Do one of the following:
 - 1. Select Clusters Cloudera Management Service .
 - 2. Select ActionsRestart.
 - On the HomeStatus tab, click  to the right of Cloudera Management Service and select Restart.
 - b. Click Restart to confirm. The Command Details window shows the progress of stopping and then starting the roles.
 - c. When Command completed with *n/n* successful subcommands appears, the task is complete. Click Close.

Installing the Parcel



Note: It is not required that the Cloudera Manager server host be part of a managed cluster and have an agent installed. Although you initially copy the CSD file to the Cloudera Manager server, the Parcel for the add-on service will not be installed on the Cloudera Manager Server host unless the host is managed by Cloudera Manager.

If you have already installed the external software onto your cluster, you can skip these steps and proceed to [Adding an Add-on Service](#) on page 22.

1. Click Parcels in the left menu. If the vendor has included the location of the repository in the CSD, the parcel should already be present and ready for downloading. If the parcel is available, skip to [step 7](#).
2. Use one of the following methods to open the parcel settings page:
 - Navigation bar
 - a. Click the parcel icon in the top navigation bar or click Hosts and click the Parcels tab.
 - b. Click the Configuration button.
 - Menu
 - a. Select AdministrationSettings.
 - b. Select CategoryParcels.
3. In the Remote Parcel Repository URLs list, click the addition symbol to open an additional row.
4. Enter the path to the repository.
5. Enter a Reason for change, and then click Save Changes to commit the changes.
6. Click Parcels in the left navigation menu. The external parcel should appear in the set of parcels available for download.
7. Download, distribute, and activate the parcel.

Adding an Add-on Service

Add the service following the procedure in [Adding a Service](#) on page 5.

Uninstalling an Add-on Service

1. Stop all instances of the service.
2. Delete the service from all clusters. If there are other services that depend on the service you are trying to delete, you must delete those services first.
3. Log on to the Cloudera Manager Server host and remove the CSD file.
4. Restart the Cloudera Manager Server:

```
service cloudera-scm-server restart
```

5. After the server has restarted, log into the Cloudera Manager Admin Console and restart the Cloudera Management Service.
6. Optionally remove the parcel.

Configuring Services to Use LZO Compression

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

After you [install the GPL Extras parcel](#), reconfigure and restart services that need to use LZO functionality. Any service that does not require the use of LZO need not be configured.

HDFS and MapReduce

1. Go to the HDFS service.
2. Click the Configuration tab.
3. Search for the `io.compression.codecs` property.
4. In the Compression Codecs property, click in the field, then click the + sign to open a new value field.
5. Add the following two codecs:
 - `com.hadoop.compression.lzo.LzoCodec`
 - `com.hadoop.compression.lzo.LzopCodec`

6. Save your configuration changes.
7. Restart HDFS.
8. Redeploy the HDFS client configuration.

Oozie

1. Go to `/var/lib/oozie` on each Oozie server and even if the LZO JAR is present, symlink the Hadoop LZO JAR:

```
/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/hadoop-lzo.jar
```

2. Restart Oozie.

HBase

Restart HBase.

Impala

Restart Impala.

Hive

Restart the Hive server.

Hive-on-Tez

1. [install the GPL Extras parcel](#).
2. Log in to Cloudera Manager and go to the Tez service.
3. Select the Configuration tab.
4. Add the following value to the Tez Additional Classpath configuration parameter:

```
/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/hadoop-lzo.jar
```

5. Append the following to the Tez Application Master Environment Settings configuration parameter:

```
:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/native
```

6. Append the following to the Tez Task Environment Settings configuration parameter:

```
:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/native
```

7. Deploy the Client Configuration (ActionsDeploy Client Configuration).
8. Go to the Hive on Tez service.
9. Restart the Hive on Tez Service (ActionsRestart).
10. Before issuing a query, go to the Tez service Configuration tab and change the value of the Codec for Compressing Intermediate Data to:

```
com.hadoop.compression.lzo.LzoCodec
```

Sqoop 1

1. Add the following entries to the Sqoop 1 Client Client Advanced Configuration Snippet (Safety Valve)
 - `HADOOP_CLASSPATH=$HADOOP_CLASSPATH:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/`
 - `JAVA_LIBRARY_PATH=$JAVA_LIBRARY_PATH:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/native`
2. Re-deploy the client configuration.