Cloudera Manager 7.13.2

Managing Clusters

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



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. 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

Accessing the Cloudera Manager Admin Console....8

Accessing the Cloudera Manager Admin Console from Data Hub clusters....8

Accessing the Cloudera Manager for Data Lake clusters using workload credentials....8

Adding and Deleting Clusters....9

About Data Context....9
Creating a Compute Cluster and Data Context....10
Adding a Cluster Using New Hosts....11
Step 1: Welcome (Add Cluster - Installation)....12
Step 2: Cluster Basics....12
Step 3: Setup Auto-TLS....12
Step 4: Specify Hosts....13
Step 5: Select Repository....14
Step 6: Select JDK....14
Step 7: Enter Login Credentials....15
Step 8: Install Agents....15
Step 9: Install Parcels....15
Step 11: Inspect Cluster....16
Adding a Cluster Using Currently Managed Hosts....16
Step 1: Welcome (Add Cluster - Installation)....16
Step 2: Cluster Basics....16
Step 3: Setup Auto-TLS....17
Step 4: Specify Hosts....18
Step 5: Select Repository....18
Step 6: Install Parcels....19
Step 8: Inspect Cluster....19
Deleting a Cluster....19
Tutorial: Using Impala, Hive and Hue with Virtual Private Clusters....19
Set Up an Environment....19
Using impala-shell and Hive....22
View HDFS directory structure of Compute clusters....27
Insert data in test_table through Spark....28
Hue in a Virtual Private Cluster Environment....30

Starting, Stopping, Refreshing, and Restarting a Cluster....32

Renaming a Cluster....33

Managing Hosts....34
Viewing Host Status....34
Adding a Host to a Cluster....35
Using the Add Hosts Wizard to Add Hosts....36
Adding a Host by Installing the Packages Using Your Own Method....42
Parcels....42
Configuring Hosts....42
Viewing Host Role Assignments....43
Host Templates....43
Creating a Host Template....43
Editing a Host Template....43
Deleting a Host Template....44
Applying a Host Template to a Host....44
Hosts Disks Overview....45
Deleting Hosts....45
Deleting a Host from Cloudera Manager....45
Removing a Host From a Cluster....45
Stopping All the Roles on a Host....46
Starting All the Roles on a Host....46
Changing Hostnames....46
Moving a Host Between Clusters....48
Configuring Upgrade Domains....49
Configuring Upgrade Domains....49
Changing the Upgrade Domain for hosts....50
Putting all Hosts in an Upgrade Domain group into Maintenance Mode....50
Specifying Racks for Hosts....51
Performing Maintenance on a Cluster Host....51
Decommissioning Hosts....52
Recommissioning Hosts....53
Tuning and Troubleshooting Host Decommissioning....54
Maintenance Mode....57
Viewing the Maintenance Mode Status of a Cluster....59

Managing Roles....60
Role Instances....60
Adding a Role Instance....60
Starting, Stopping, and Restarting Role Instances....61
Decommissioning Role Instances....62
Recommissioning Role Instances....63
Deleting Role Instances....63
Deleting Role Instances from a Host Template....64
Configuring Roles to Use a Custom Garbage Collection Parameter....64
Role Groups....64
Creating a Role Group....65
Managing Role Groups....65

Default User Roles....66

Backing up Cloudera Manager databases....67
Backing Up PostgreSQL Databases....67
Backing Up MariaDB Databases....68

Backing Up MySQL Databases 68
Backing Up Oracle Databases 68

Core Settings Service 69
Configuration parameters migrated to Core Settings Service 70

Performance Management 71
Optimizing Performance in Cloudera Runtime 71
Disabling Transparent Hugepages (THP) 72
Setting the vm.swappiness Linux Kernel Parameter 73
Improving Performance in Shuffle Handler and IFile Reader 74
Tips and Best Practices for Jobs 75
Decrease Reserve Space 75
Choosing and Configuring Data Compression 76

Managing Cloudera Manager 76
Automatic Logout 76
Starting, Stopping, and Restarting the Cloudera Manager Server 77
Configuring Cloudera Manager 77
Configuring Cloudera Manager Server Ports 79
Configuring Network Settings for a Proxy Server 79
Moving the Cloudera Manager Server to a New Host 79
Migrating Cloudera Manager Management Services to Another Host 80
Migrating from the Cloudera Manager Embedded PostgreSQL Database Server to an External PostgreSQL Database 83
Step 1: Identify Roles that Use the Embedded Database Server 84
Step 2: Migrate Databases from the Embedded Database Server to the External PostgreSQL Database Server 86
Migrating from the Cloudera Manager External PostgreSQL Database Server to a MySQL/Oracle Database Server 90
Prerequisites 90
Migrate from the Cloudera Manager External PostgreSQL Database Server to a MySQL/Oracle Database Server 91
Managing Cloudera Manager Server Logs 94
Cloudera Manager Agents 97
Starting, Stopping, and Restarting Cloudera Manager Agents 98
Configuring Cloudera Manager Agents 99
Managing the Cloudera Manager Agent Logs 101
Overview of Parcels 102
Advantages of Parcels 102
Parcel Life Cycle 103
Parcel Locations 104
Managing Parcels 104
Viewing Parcel Usage 108
Parcel Configuration Settings 109
Managing Licenses and Cluster Registration 111
Accessing the License Page 111
Upgrading from a Cloudera Private Cloud Base Trial to Cloudera Private Cloud Base 112
Renewing a License 112
Mandatory Cluster Registration Policy 112
Registering and Deregistering Clusters in Cloudera Manager 114
Ending a Cloudera Private Cloud Base Trial 116
FAQ for Cluster Registration and V2 Licenses 116

Cloudera Manager User Roles ... 118
Displaying Your Roles ... 119
Default User Roles ... 119
User Roles with Privileges for a Cluster ... 119
Adding a User Role for a Specific Cluster ... 120
Mapping External Authentication to a Role ... 121
Assigning Users to a Role ... 122
Removing a User or External Mapping from a User Role ... 122
Removing a Role ... 122
Removing the Full Administrator User Role ... 122
Other Cloudera Manager Tasks and Settings ... 123
Providing full administrative permissions to the users on a temporary basis ... 124

Cloudera Management Service ... 125
Starting the Cloudera Management Service ... 127
Stopping the Cloudera Management Service ... 127
Restarting the Cloudera Management Service ... 127
Starting and Stopping Cloudera Management Service Roles ... 128
Configuring Management Service Database Limits ... 128

Configuring Cloudera Manager for High Availability ... 129
Prerequisites for configuring Cloudera Manager for high availability ... 130
User impacts and limitations for Cloudera Manager in high availability mode ... 131
Installing Cloudera Manager with High Availability ... 135
Sample Configuration Files for HAProxy ... 137

Securing sensitive information using a Secure Credential Storage Provider ... 140
Configuring a Secure Credential Storage Provider for Cloudera Manager ... 140
Disabling the Credential Storage Provider ... 142
Managing Credential Storage Provider ... 142

Resource Management ... 143
Static Service Pools ... 144
Enabling and Configuring Static Service Pools ... 146
Describing different columns in a Static Service Pool Wizard ... 149
Disabling Static Service Pools ... 150
Linux Control Groups (cgroups) ... 151
Enabling Resource Management with Control Groups ... 151
Configuring Resource Parameters ... 152
Configuring Custom Cgroups ... 156
Migrating from Cgroup v1 to Cgroup v2 ... 157
Data Storage for Monitoring Data ... 158
Configuring Service Monitor Data Storage ... 158
Configuring Host Monitor Data Storage ... 158
Viewing Host and Service Monitor Data Storage ... 159
Data Granularity and Time-Series Metric Data ... 159
Moving Monitoring Data on an Active Cluster ... 159
Host Monitor and Service Monitor Memory Configuration ... 160

Accessing Storage Using Microsoft ADLS Gen 2.....162
Configuring OAuth in Data Hub.....162
Configuring OAuth with core-site.xml.....162
Configuring OAuth with the Hadoop CredentialProvider.....163
Configuring Built-in TLS Acceleration.....164
Importing Data into Microsoft Azure Data Lake Store (Gen1 and Gen2) Using Sqoop.....164
Prerequisites.....165
Authentication.....165
Sqoop Import into ADLS.....165

Configuration Change Audit API.....166

Accessing the Cloudera Manager Admin Console

How to open and log in to the Cloudera Manager Admin Console.

About this task
To access the Cloudera Manager Admin Console:

Procedure

1. Open the Cloudera Management Console in a Web browser using the following URL:

   ```
   http://<CLOUDERA MANAGER SERVER URL>:7180
   ```

2. Enter your user name.
3. Enter your password.
4. Click the Sign In button.

Results
The Cloudera Manager Admin Console opens.

Accessing the Cloudera Manager Admin Console from Data Hub clusters

After you create a Data Hub cluster using the Cloudera Management Console, you can access the Cloudera Manager Admin Console to manage, configure, and monitor the cluster and its Cloudera Runtime services.

About this task
To access the Cloudera Manager Admin Console:

Procedure

1. Open the CDP web interface at https://console.cdp.cloudera.com and log in by using your corporate credentials or other credentials that you received from your CDP administrator.
2. Click Data Hub Clusters.
3. Click the name of the cluster you want to manage.
   The cluster details page displays.
4. Click the URL for Cloudera Manager.

Results
The Cloudera Manager Admin Console opens in a new browser tab. You do not need to login to the Cloudera Manager Admin Console.

Accessing the Cloudera Manager for Data Lake clusters using workload credentials

When you access Cloudera Manager from a Data Lake cluster through the Management Console, the system authenticates you using SSO. If you want to use credentials instead of SSO login to log in to the Cloudera manager then you can use workload credentials. To log in to the Cloudera Manager using workload credentials, you must use a different URL.

About this task
To log in to the Cloudera Manager using workload credentials:

Before you begin
You must have the IP address of the host on which Cloudera Manager is running.

Procedure
1. Open a web browser and specify the following URL in the address bar:

```
https://[***CM-HOST-IP-ADDRESS***]/clouderamanager/
```



Important: You must use the Cloudera Manager server IP address in the above URL.

2. Enter your workload user name and password.
3. Click Sign In.

Results
The Cloudera Manager Admin Console opens for the Data Lake cluster.

Adding and Deleting Clusters

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

Cloudera Manager can manage multiple clusters. The clusters do not need to run the same major version of CDH or Cloudera Runtime.

About Data Context

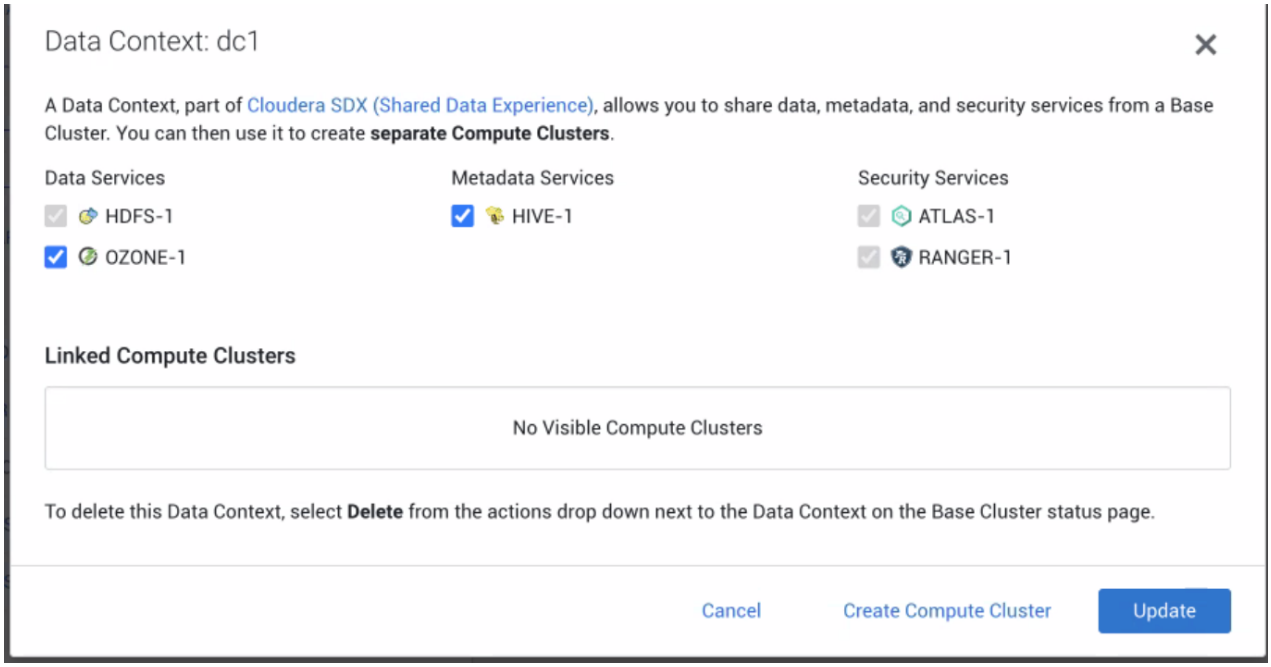
Data Contexts in Cloudera Manager are used to access data in Cloudera Private Cloud Base environment. In other words, a Data Context helps you to share the services from one cluster to another.

A Cloudera Private Cloud Base cluster can have one or more Compute clusters. A Compute cluster employs a Data Context to connect to the Base cluster for accessing the data and metadata in the Base cluster. The context itself is merely a logical entity where there is no specific deployment or cluster running activities.

When the Data Context is created, you can view the list of available services within the Base cluster that you can share with compute cluster: this list is a subset of the services in the cluster, as not all services have the ability to be shared through Data Context.



Important: Previously, when clicking on a Data Context link in a Base cluster, only the content of the Data Context was displayed. From Cloudera Runtime 7.1.9 release, you can modify the service selection. You can add or remove certain services to the Data Context.



Note the following:

- A Data Context can only be used to connect a Base cluster with one or more compute clusters. A Data Context cannot be used to connect two Base clusters.
- You can connect your compute cluster to a Base cluster only if the Data Context is made available. Later you can use the services available in the Base cluster because the Data Context provides the metadata to connect the Virtual Private Cluster to the Base cluster.
- The Compute cluster needs configuration files and some information to be able to communicate with a Base cluster. All configuration files are managed by Cloudera Manager automatically, and will be updated as required if any service in the Data Context has changed their configurations.

Creating a Compute Cluster and Data Context

Compute cluster needs Data Context to communicate with the Base cluster and you must create one.

Before you begin

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

To create a Compute cluster, you must have a Base cluster. The Base cluster contains data that is used by the Compute cluster.

To create a Compute cluster:

Procedure

1. On the Cloudera Manager home page, click Clusters Add Cluster
   The Add Cluster Welcome page displays.
2. Click Continue. .
   The Cluster Basics page displays
3. Select Compute cluster.
4. If you already have a Data Context defined, select it from the drop-down list.
5. To create a new Data Context:
   a) Select Create Data Context from the drop-down list.

The Create Data Context dialog box displays.

b) Enter a unique name for the Data Context.

c) Select the Base cluster from the drop-down list.

d) Select the services you want to expose in the Data Context. You can choose from the following:

- HDFS (required)
- Hive Metadata Service
- Atlas
- Ranger
- Ozone

e) Click Create.
The Cluster Basics page displays your selections.

f) Click Continue.

6. Continue with the next steps in the Add Cluster Wizard to specify hosts and credentials, and install the Agent and CDH software.

   The Select Repository screen examines the CDH version of the base cluster and recommend a supported version. Cloudera recommends that your Base and Compute clusters each run the same version of CDH. The Add Cluster Wizard offers the option to choose other versions, but these combinations have not been tested and are not supported for production use.

7. On the Select Services screen, choose any of the pre-configured combinations of services listed on this page, or you can select Custom Services and choose the services you want to install.

   Service combinations for Compute Clusters:

   The following services can be installed on a Compute cluster:

   - Hive Execution Service (This service supplies the HiveServer2 role only.)
   - Hue
   - Kafka
   - Spark
   - Oozie (only when Hue is available, and is a requirement for Hue)
   - YARN
   - HDFS
   - Stub DFS (Stub DFS replaces Core Settings and requires the Storage Operations role.)

8. If you have enabled Kerberos authentication on the Base cluster, you must also enable Kerberos on the Compute cluster.

9. If you don't select a service from Step 7 but if you want to use integrations related to these services (such as Hive, Spark, Kafka, etc.) on the Compute cluster, then you need to add all the necessary gateway roles to use the connections properly.

Adding a Cluster Using New Hosts

Use the Add Cluster - Installation wizard to create a regular or computer cluster using new hosts.

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

On the Cloudera Manager Home page, click the Add drop-down button at the top right, or the Clusters drop-down button at the top left, and then click Add Cluster. This launches the Add Cluster - Installation wizard, which allows you to create either a regular cluster or a compute cluster.

You can also launch the wizard by selecting Add Compute Cluster from the drop-down menu next to the cluster name. Launching the wizard from there skips the Welcome page and restricts the wizard to creating only a compute cluster.

Step 1: Welcome (Add Cluster - Installation)

The Welcome page of the Add Cluster - Installation wizard provides a brief overview of the installation and configuration procedure, as well as some links to relevant documentation.

If you launched the wizard using the Add Compute Cluster option, this page is not displayed.

Click Continue to proceed with the installation.

Step 2: Cluster Basics

The Cluster Basics page allows you to specify the Cluster Name and select the Cluster Type.

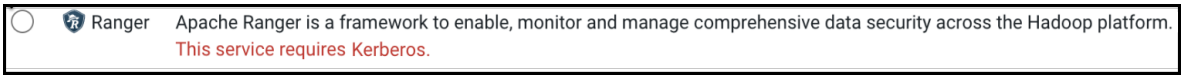
There are two cluster types:

- Regular Cluster: A Regular Cluster contains storage nodes, compute nodes, and other services such as metadata and security collocated in a single cluster.
- Compute Cluster: A Compute Cluster consists of only compute nodes. To connect to existing storage, metadata or security services, you must first choose or create a Data Context on a Base Cluster.

Important:

In case you deploy a scratch cluster that is not yet kerberised, you cannot select any service that requires Kerberos as a prerequisite including Ranger, RangerKMS, Atlas, and Knox.

Each of these services carries such a note near it (see example for Ranger):



Selecting either of these services to be installed without having a Kerberos based cluster leads to a non-functioning cluster. It is advised to install these services separately only after enabling Kerberos on the cluster. For more information, see Set up a KDC.

Note that if you do not add Kerberos to the cluster at the beginning of the wizard (for example, setting Cloudera Manager as Key Distribution Center), you need to selct only Customer Services option and pick the non-kerberos services you want to deploy. Post enabling Kerberos, you need to repeat the wizard and complete the Kerberos based services.

If you are performing a new installation, Regular Cluster is the only option. You cannot add a compute cluster if you do not have an existing base cluster.

For more information on regular and compute clusters, and data contexts, see *Virtual Private Clusters and Cloudera SDX*.

If you are adding a compute cluster to an existing base cluster, click Choose Data Context... to create or select a data context.

After selecting a cluster type and data context (if applicable), enter a cluster name and then click Continue.

After a cluster is created, you can change the cluster display name using the Rename Cluster operation, but the internal name remains same. For more information, see *Renaming a cluster*.

Related Information

Renaming a Cluster

Step 3: Setup Auto-TLS

The Setup Auto-TLS page provides instructions for initializing the certificate manager for auto-TLS if you have not done so already.

Important: Auto-TLS is only available with an Enterprise license.

If you already initialized the certificate manager in *Step 3: Install Cloudera Manager Server*, the wizard displays a message indicating that auto-TLS has been initialized. Click Continue to proceed with the installation.

If you have not already initialized the certificate manager, and you want to enable auto-TLS, follow the instructions provided on the page before continuing. When you reload the page as instructed, you are redirected to https://<*SERVER_HOST*>:7183, and a security warning is displayed. You might need to indicate that you trust the certificate, or click to proceed to the Cloudera Manager Server host. You might also be required to log in again and re-complete the previous steps in the wizard.

For more information, see *Configuring TLS Encryption for Cloudera Manager Using Auto-TLS*.

If you do not want to enable auto-TLS at this time, click Continue to proceed.

Related Information

Configuring TLS Encryption for Cloudera Manager Using Auto-TLS

Step 4: Specify Hosts

Cloudera Manager can automatically discover hosts on which to install Cloudera Runtime and managed services. This section covers the procedure for new hosts only.

Before you begin

Choose which hosts will run Cloudera Runtime and other managed services.

Procedure

1. To enable Cloudera Manager to automatically discover hosts on which to install CDH and managed services, enter the cluster hostnames or IP addresses in the Hostnames field. You can specify hostname and IP address ranges as follows:

Expansion Range	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].example.com	host1.example.com, host2.example.com, host3.example.com
host[07-10].example.com	host07.example.com, host08.example.com, host09.example.com, host10.example.com



Important: Unqualified hostnames (short names) must be unique in a Cloudera Manager instance. For example, you cannot have both *HOST01.EXAMPLE.COM* and *HOST01.STANDBY.EXAMPLE.COM* managed by the same Cloudera Manager Server.

You can specify multiple addresses and address ranges by separating them with commas, semicolons, tabs, or blank spaces, or by placing them on separate lines. Use this technique to make more specific searches instead of searching overly wide ranges. Only scans that reach hosts running SSH will be selected for inclusion in your cluster by default. You can enter an address range that spans over unused addresses and then clear the nonexistent hosts later in the procedure, but wider ranges require more time to scan.

2. Click Search. If there are a large number of hosts on your cluster, wait a few moments to allow them to be discovered and shown in the wizard. If the search is taking too long, you can stop the scan by clicking Abort Scan. You can modify the search pattern and repeat the search as many times as you need until you see all of the expected hosts.



Note: Cloudera Manager scans hosts by checking for network connectivity. If there are some hosts where you want to install services that are not shown in the list, make sure you have network connectivity between the Cloudera Manager Server host and those hosts, and that firewalls and SE Linux are not blocking access.

3. Verify that the number of hosts shown matches the number of hosts where you want to install services. Clear host entries that do not exist or where you do not want to install services.
4. Click Continue.
   The Select Repository screen displays.

Step 5: Select Repository

The Select Repository page allows you to specify repositories for Cloudera Manager Agent and CDH and other software.

Procedure

1. In the Cloudera Manager Agent section, select either Public Cloudera Repository or Custom Repository for the Cloudera Manager Agent software.
2. If you select Custom Repository, do not include the operating system-specific paths in the URL. For instructions on setting up a custom repository, see *Configuring a Local Package Repository*.
3. In the CDH and other software section, select the repository type to use for the installation. In the Install Method section select one of the following:
   - Use Parcels (Recommended)

     A parcel is a binary distribution format containing the program files, along with additional metadata used by Cloudera Manager. Parcels are required for rolling upgrades.
   - Use Packages

     A package is a standard binary distribution format that contains compiled code and meta-information such as a package description, version, and dependencies. Packages are installed using your operating system package manager.
4. Select the version of Cloudera Runtime to install. For compute clusters using parcels, the supported Cloudera Runtime versions display (Supported) next to the parcel name. For compute clusters using packages, you must make sure that you have installed a supported Cloudera Runtime version on all compute cluster hosts.
   a) If you selected Use Parcels and you do not see the version you want to install, click the More Options button to add the repository URL for your version. Repository URLs for Cloudera Runtime 7.x are documented in the Release Guide. After adding the repository, click Save Changes and wait a few seconds for the version to appear. If your Cloudera Manager host uses an HTTP proxy, click the Proxy Settings button to configure your proxy.

   Note: Cloudera Manager only displays Cloudera Runtime versions it can support. If an available Cloudera Runtime version is too new for your Cloudera Manager version, it is not displayed.

   b) If you selected Use Packages, and the version you want to install is not listed, you can select Custom Repository to specify a repository that contains the desired version. Repository URLs for Cloudera Runtime 7.x are documented in the Release Guide.
5. If you selected Use Parcels, specify any Additional Parcels you want to install.
6. Click Continue.

   The Accept JDK License page displays.

Step 6: Select JDK

Cloudera Manager can automatically install a JDK on cluster hosts, or you can choose to install the JDK manually.



Note: If you want to install OpenJDK 11* or OpenJDK 17*, you must install it manually on all hosts before continuing.

1. Choose one of the following options:
   - Manually manage JDK

     If you select this option, you must ensure that a supported JDK is already installed on all hosts. You will need to manage installing the unlimited strength JCE policy file, if necessary. If you are installing OpenJDK 11* or OpenJDK 17*, you must select this option.

* Azul OpenJDK, OpenJDK 8, OpenJDK 11, and OpenJDK 17 are TCK certified for CDP.

- Install a Cloudera-provided version of OpenJDK 8

  Cloudera Manager installs a supported version of OpenJDK 8* on all hosts.
- Install a system-provided version of OpenJDK

  Cloudera Manager installs the default version of OpenJDK 8* provided by the operating system.

2. Click Continue.

Related Information

Cloudera Support Matrix

Step 7: Enter Login Credentials

Enter the required login credentials and choose the method of authentication.

Procedure

1. Select root for the root account, or select Another user and enter the username for an account that has password-less sudo privileges.
2. Select an authentication method:
   - If you choose password authentication, enter and confirm the password.
   - If you choose public-key authentication, provide a passphrase and path to the required key files.

   You can modify the default SSH port if necessary.
3. Specify the maximum number of host installations to run at once. The default and recommended value is 10. You can adjust this based on your network capacity.
4. Click Continue.
   The Install Agents page displays.

Step 8: Install Agents

The Install Agents page displays the progress of the installation.

You can click on the Details link for any host to view the installation log. If the installation is stalled, you can click the Abort Installation button to cancel the installation and then view the installation logs to troubleshoot the problem.

If the installation fails on any hosts, you can click the Retry Failed Hosts to retry all failed hosts, or you can click the Retry link on a specific host.

If you selected the option to manually install agents, see *Manually Install Cloudera Manager Agent Packages* for the procedure and then continue with the next steps on this page.

After installing the Cloudera Manager Agent on all hosts, click Continue.

If you are using parcels, the Install Parcels page displays. If you chose to install using packages, the Detecting CDH Versions page displays.

Step 9: Install Parcels

If you selected parcels for the installation method, the Install Parcels page reports the installation progress of the parcels you selected earlier.

After the parcels are downloaded, progress bars appear representing each cluster host. You can click on an individual progress bar for details about that host.

After the installation is complete, click Continue.

The Inspect Cluster page displays.

Step 11: Inspect Cluster

The Inspect Cluster page provides a tool for inspecting network performance as well as the Host Inspector to search for common configuration problems. Cloudera recommends that you run the inspectors sequentially.

About this task

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

Procedure

1. Run the Inspect Network Performance tool. You can click Advanced Options to customize some ping parameters.
2. After the network inspector completes, click Show Inspector Results to view the results in a new tab.
3. Address any reported issues, and click Run Again (if applicable).
4. Click Inspect Hosts to run the Host Inspector utility.
5. After the host inspector completes, click Show Inspector Results to view the results in a new tab.
6. Address any reported issues, and click Run Again (if applicable).
7. If the reported issues cannot be resolved in a timely manner, and you want to abandon the cluster creation wizard to address them, select the radio button labeled Quit the wizard and Cloudera Manager will delete the temporarily created cluster and then click Continue. Otherwise, after addressing any identified problems, select the radio button labeled I understand the risks, let me continue with cluster creation, and then click Continue.

Results

This completes the Add Cluster - Installation wizard and launches the Add Cluster - Configuration wizard. For further instructions, see *Set up a Cluster Using the Wizard* in the Installation guide.

Adding a Cluster Using Currently Managed Hosts

This section covers the procedure for creating a cluster from existing managed hosts only. For instructions using new (currently unmanaged) hosts, see *Adding a Cluster Using New Hosts*.

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

Before continuing, make sure that the managed hosts have the desired CDH or Cloudera Runtime version packages pre-installed.

On the Cloudera Manager Home page, click the Add drop-down button at the top right, or the Clusters drop-down button at the top left, and then click Add Cluster. This launches the Add Cluster - Installation wizard, which allows you to create either a regular cluster or a compute cluster.

You can also launch the wizard by selecting Add Compute Cluster from the drop-down menu next to the cluster name. Launching the wizard from there skips the Welcome page and restricts the wizard to creating only a compute cluster.

The following sections guide you through each page of the wizard.

Step 1: Welcome (Add Cluster - Installation)

The Welcome page of the Add Cluster - Installation wizard provides a brief overview of the installation and configuration procedure, as well as some links to relevant documentation.

If you launched the wizard using the Add Compute Cluster option, this page is not displayed.

Click Continue to proceed with the installation.

Step 2: Cluster Basics

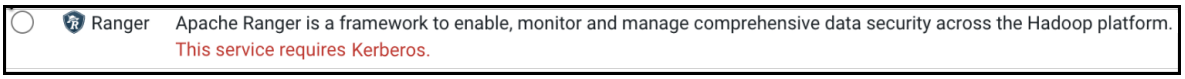
The Cluster Basics page allows you to specify the Cluster Name and select the Cluster Type:

- Regular Cluster: A Regular Cluster contains storage nodes, compute nodes, and other services such as metadata and security collocated in a single cluster.
- Compute Cluster: A Compute Cluster consists of only compute nodes. To connect to existing storage, metadata or security services, you must first choose or create a Data Context on a Base Cluster.

Important:

In case you deploy a scratch cluster that is not yet kerberised, you cannot select any service that requires Kerberos as a prerequisite including Ranger, RangerKMS, Atlas, and Knox.

Each of these services carries such a note near it (see example for Ranger):



Selecting either of these services to be installed without having a Kerberos based cluster leads to a non-functioning cluster. It is advised to install these services separately only after enabling Kerberos on the cluster. For more information, see Set up a KDC.

Note that if you do not add Kerberos to the cluster at the beginning of the wizard (for example, setting Cloudera Manager as Key Distribution Center), you need to selct only Customer Services option and pick the non-kerberos services you want to deploy. Post enabling Kerberos, you need to repeat the wizard and complete the Kerberos based services.

For more information on regular and compute clusters, and data contexts, see *Virtual Private Clusters and Cloudera SDX*.

If you are adding a compute cluster to an existing base cluster, click Choose Data Context... to create or select a Data Context.

After selecting a cluster type and data context (if applicable), enter a cluster name and then click Continue.

After a cluster is created, you can change the cluster display name using the Rename Cluster operation, but the internal name remains same. For more information, see *Renaming a cluster*.

Related Information

Renaming a Cluster

Step 3: Setup Auto-TLS

The Setup Auto-TLS page provides instructions for initializing the certificate manager for auto-TLS if you have not done so already.

Important: Auto-TLS is only available with an Enterprise license.

If you already initialized the certificate manager in *Step 3: Install Cloudera Manager Server*, the wizard displays a message indicating that auto-TLS has been initialized. Click Continue to proceed with the installation.

If you have not already initialized the certificate manager, and you want to enable auto-TLS, follow the instructions provided on the page before continuing. When you reload the page as instructed, you are redirected to https://<*SERVER_HOST*>:7183, and a security warning is displayed. You might need to indicate that you trust the certificate, or click to proceed to the Cloudera Manager Server host. You might also be required to log in again and re-complete the previous steps in the wizard.

For more information, see *Configuring TLS Encryption for Cloudera Manager Using Auto-TLS*.

If you do not want to enable auto-TLS at this time, click Continue to proceed.

Related Information

Configuring TLS Encryption for Cloudera Manager Using Auto-TLS

Step 4: Specify Hosts

This section covers the procedure for creating a cluster from existing managed hosts only. For instructions using new (currently unmanaged) hosts, see *Adding a Cluster Using New Hosts*.

Select the hosts for your cluster by clicking the Currently Managed Hosts tab. This tab does not appear if you have no unassigned managed hosts. You cannot select a mixture of new hosts and currently managed hosts.

If you are installing CDH and other services using packages instead of parcels, make sure that you have manually installed the CDH packages on each host before continuing.

Select the hosts you want to add to the cluster, and then click Continue.

Step 5: Select Repository

The Select Repository page allows you to specify repositories for Cloudera Manager Agent and CDH and other software. If you are installing on currently managed hosts, the Cloudera Manager Agent section is not displayed.

About this task



Important: You cannot install software using both parcels and packages in the same cluster.

Procedure

1. In the Cloudera Manager Agent section, select either Public Cloudera Repository or Custom Repository for the Cloudera Manager Agent software. If you select Custom Repository, do not include the operating system-specific paths in the URL. For instructions on setting up a custom repository, see *Configuring a Local Package Repository*.
2. In the CDH and other software section, select the repository type to use for the installation. In the Install Method section select one of the following:
   - Use Parcels (Recommended)

     A parcel is a binary distribution format containing the program files, along with additional metadata used by Cloudera Manager. Parcels are required for rolling upgrades. For more information, see Overview of Parcels on page 102.
   - Use Packages

     A package is a standard binary distribution format that contains compiled code and meta-information such as a package description, version, and dependencies. Packages are installed using your operating system package manager.

   If you select Use Packages, make sure that you have manually installed the CDH packages on each host before continuing.
3. Select the version of CDH or Runtime to install. For compute clusters using parcels, the supported versions display (Supported) next to the parcel name. For compute clusters using packages, you must make sure that you have installed a supported CDH or Runtime version on all compute cluster hosts.

   If you selected Use Parcels and you do not see the version you want to install, click the More Options button to add the repository URL for your version. Repository URLs for CDH version 6 and Cloudera Runtime are documented in the Release Guide. After adding the repository, click Save Changes and wait a few seconds for the version to appear. If your Cloudera Manager host uses an HTTP proxy, click the Proxy Settings button to configure your proxy.

   Note: Cloudera Manager only displays CDH versions it can support. If an available CDH version is too new for your Cloudera Manager version, it is not displayed.
4. If you selected Use Parcels, specify any Additional Parcels you want to install. If you are installing CDH 6, do not select the KAFKA, KUDU, or SPARK parcels, because they are included in CDH 6.
5. Click Continue.

If you are using parcels, the Install Parcels page displays. If you chose to install using packages, the Detecting CDH Versions page displays.

Step 6: Install Parcels

If you selected parcels for the installation method, the Install Parcels page reports the installation progress of the parcels you selected earlier.

After the parcels are downloaded, progress bars appear representing each cluster host. You can click on an individual progress bar for details about that host.

After the installation is complete, click Continue.

The Inspect Cluster page displays.

Step 8: Inspect Cluster

The Inspect Cluster page provides a tool for inspecting network performance as well as the Host Inspector to search for common configuration problems. Cloudera recommends that you run the inspectors sequentially.

Procedure

1. Run the Inspect Network Performance tool. You can click Advanced Options to customize some ping parameters.
2. After the network inspector completes, click Show Inspector Results to view the results in a new tab.
3. Address any reported issues, and click Run Again (if applicable).
4. Click Inspect Hosts to run the Host Inspector utility.
5. After the host inspector completes, click Show Inspector Results to view the results in a new tab.
6. Address any reported issues, and click Run Again (if applicable).
7. If the reported issues cannot be resolved in a timely manner, and you want to abandon the cluster creation wizard to address them, select the radio button labeled Quit the wizard and Cloudera Manager will delete the temporarily created cluster and then click Continue. Otherwise, after addressing any identified problems, select the radio button labeled I understand the risks, let me continue with cluster creation, and then click Continue.

Results

This completes the Add Cluster - Installation wizard and launches the Add Cluster - Configuration wizard. For further instructions, see *Set up a Cluster Using the Wizard* in the Installation guide.

Deleting a Cluster

Steps to delete a cluster.

About this task

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

Procedure

1. Stop the cluster.
2. On the HomeStatus tab, click the drop-down arrow to the right of the cluster name and select Delete.

Tutorial: Using Impala, Hive and Hue with Virtual Private Clusters

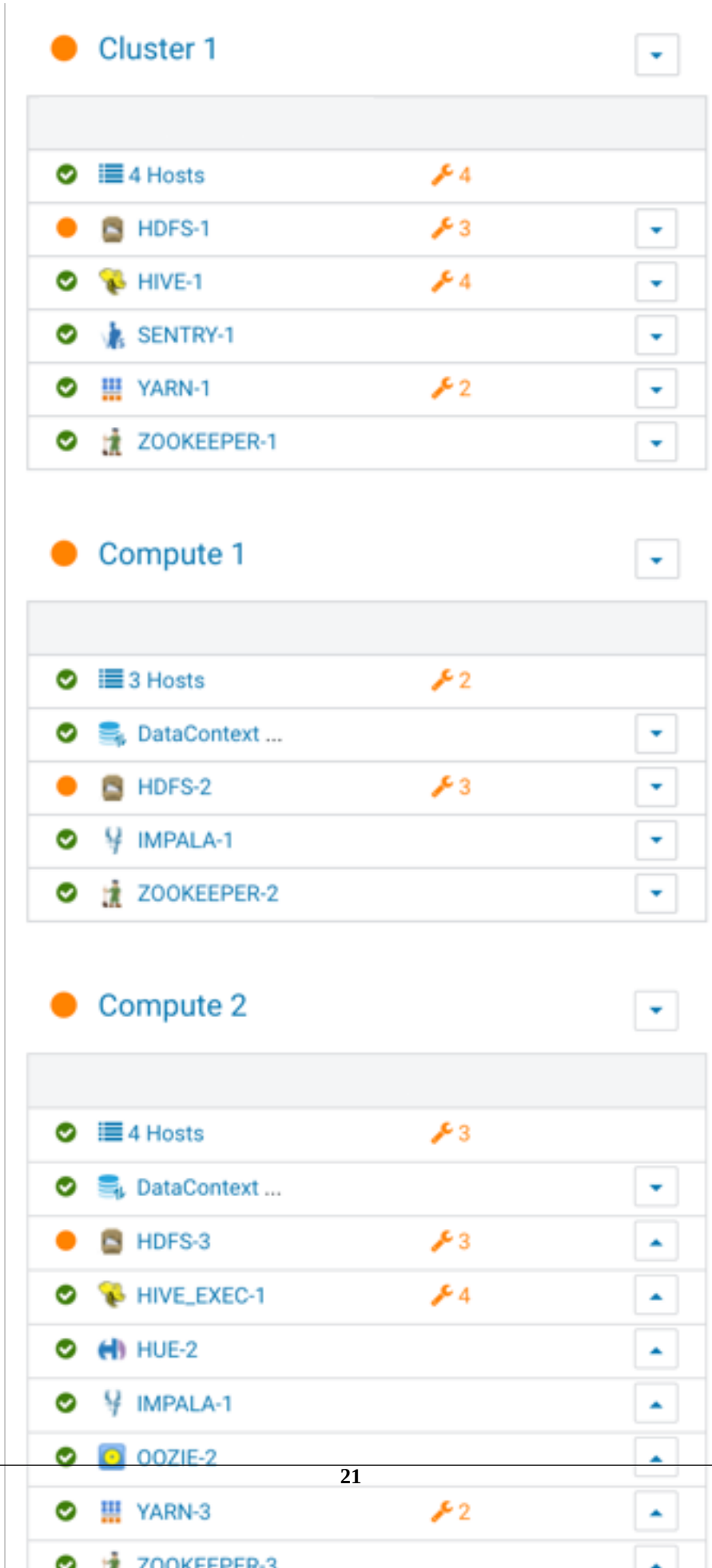
A tutorial that demonstrates how to access data in a Base cluster from a Compute cluster.

Set Up an Environment

Set up an environment to run this tutorial.

Set up your environment with Compute and Base clusters as follows:

1. Create clusters where the Cloudera Manager and Cloudera Runtime version match, for example both are Cloudera Runtime 7.1.1. The clusters must use Kerberos and TLS.
2. If Base cluster has Sentry, then make sure the user executing cross cluster queries is added to the correct role that has all the necessary privileges to create/insert data into tables. (more in workflow #3).
3. Configure a Regular cluster called Cluster 1 to be used as a Base cluster. This cluster must have high availability enabled.
4. Create two Compute clusters called Compute 1 and Compute 2.
5. Configure services for the three clusters as shown below:


Cluster 1
4 Hosts
4
HDFS-1
3
HIVE-1
4
SENTRY-1
YARN-1
2
ZOOKEEPER-1
Compute 1
3 Hosts
2
DataContext ...
HDFS-2
3
IMPALA-1
ZOOKEEPER-2
Compute 2
4 Hosts
3
DataContext ...
HDFS-3
3
HIVE_EXEC-1
4
HUE-2
IMPALA-1
OOZIE-2
YARN-3
2

Using impala-shell and Hive

Learn how to use Impala to create tables, insert data, access data, and modify data in a Virtual Private Cluster.

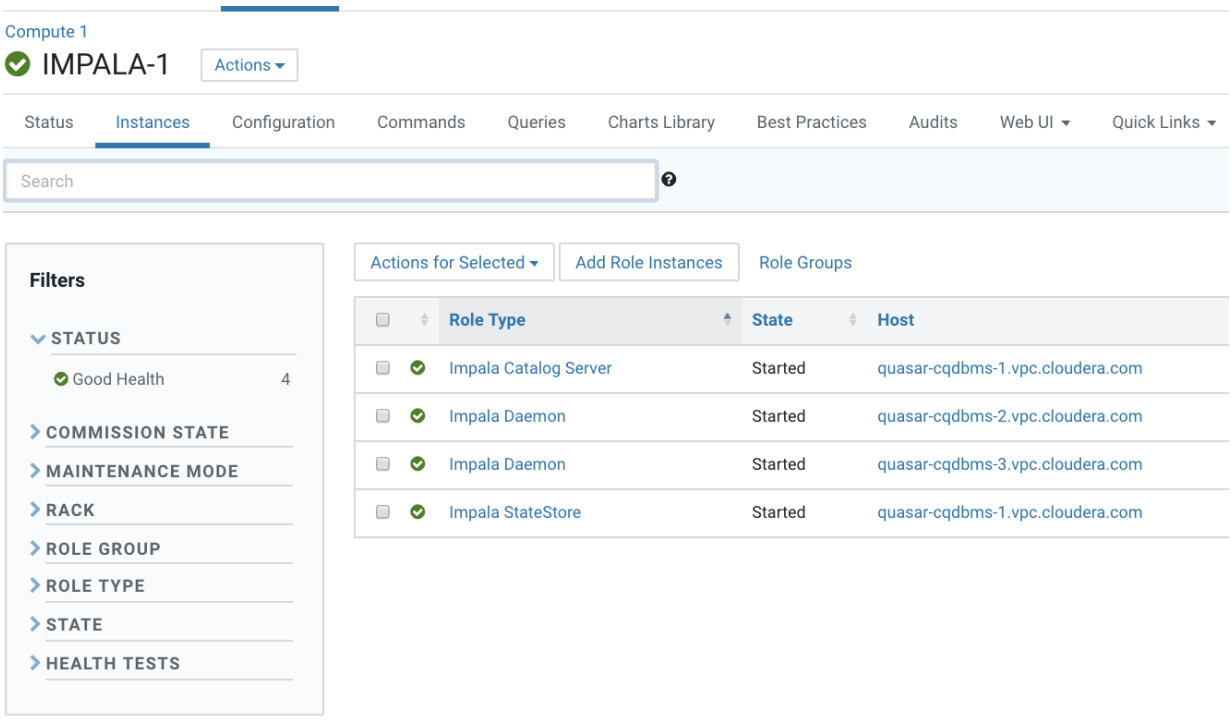
This workflow describes how to create a table using Impala, how to insert sample data on Compute cluster 1, and how to access and modify the data using beeline from Compute cluster 2.

In a Virtual Private Cluster environment, Hue and the impala-shell can be used to setup databases, tables, insert and retrieve data using queries. This workflow focuses on running a few queries using impala-shell command line tool.

This tutorial uses a kerberized environment with TLS, so you must kinit the user first. If you want to add a new user, see Enabling Kerberos Authentication for CDP for documentation to create and add the user to the Kerberos principal and the required Linux groups.

1. Identify a host running Impala Daemon to launch impala-shell using this breadcrumb. In the Cloudera Manager Admin Console, go to Cloudera Manager Compute Cluster 1 IMPALA-1 Instances .
2. Note the hostname of a host that is running the Impala Daemon and open an ssh session to that host.

```
ssh <hostname>
```



3. Kinit the user (because this is a Kerberized environment):

```
kinit -kt <keytab path of authorized user> <authorized user>
```

4. Start the Impala-shell:

   In the Cloudera Manager Admin Console, go to the Impala service and click the Status tab.

   The Impala Shell command displays in the upper-left. Copy that string and use it as the command to open Impala shell. For example:

```
impala-shell -i vpc_host-cqdbms-2.tut.myco.com:25003 -d default -k --ssl
 --ca_cert=/etc/cdep-ssl-conf/CA_STANDARD/truststore.pem
```

The command uses the following form:

```
impala-shell -i <IMPALA DAEMON HOST URL>:25003 -d default -k --ssl --ca_
cert=<PATH TO TRUSTSTORE>/truststore.pem
```

```
Query: use `default`
```

```
[vpc_host-cqdbms-2.tut.myco.com:25003] default>
```

5. Verify that impala-shell is in the connected status.
6. Create a database and a table:

```
[vpc_host-cqdbms-2.tut.myco.com:25003] default> create database test_data;
```

```
Query: create database test_data
+----------------------------+
| summary                    |
+----------------------------+
| Database has been created. |
+----------------------------+
```

```
[vpc_host-cqdbms-2.tut.myco.com:25003] default> show databases;
Query: show databases
+------------------+-------------------------------------------------+
| name             | comment                                         |
+------------------+-------------------------------------------------+
| _impala_builtins | System database for Impala builtin functions |
| default          | Default Hive database                           |
| test_data        |                                                 |
+------------------+-------------------------------------------------+
```

```
[vpc_host-cqdbms-2.tut.myco.com:25003] default> use test_data;
```

```
[vpc_host-cqdbms-2.tut.myco.com:25003] test_data> create table test_table
 (year int, winner string);
+-------------------------+
| summary                 |
+-------------------------+
| Table has been created. |
+-------------------------+
Fetched 1 row(s) in 0.47s
```

7. Verify that the table has been created on the Base cluster HDFS

```
[vpc_host-cqdbms-2.tut.myco.com:25003] test_data> show create table test
_table;
```

```
Query: show create table test_table
+----------------------------------------------------------------------+
| result                                                               |
+----------------------------------------------------------------------+
| CREATE TABLE test_data.test_table (                                  |
|   year INT,                                                          |
|   winner STRING                                                      |
| )                                                                    |
| STORED AS TEXTFILE                                                   |
| LOCATION 'hdfs://ns1/user/hive/warehouse/test_data.db/test_table' |
```

```
|                                                                    |
+--------------------------------------------------------------------+
```

8. Insert test data into the table:

```
[vpc_host-cqdbms-2.tut.myco.com:25003] test_data> insert into table test
_table values (2018, 'France'), (2014, 'Germany'), (2010, 'Spain'), (2006,
 'Italy');
```

```
Query: insert into table test_table values (2018, 'France'), (2014, 'Ger
many'), (2010, 'Spain'), (2006, 'Italy')
Query progress can be monitored at: https://vpc_host-cqdbms-2.tut.myco.
com:25000/query_plan?query_id=334fc3bd7e421cce:540f171500000000
```

```
[vpc_host-cqdbms-2.tut.myco.com:25003] test_data> select * from test_tab
le;
Query progress can be monitored at: https://vpc_host-cqdbms-2.tut.myco.c
om:25000/query_plan?query_id=be4c4c177520a5fd:7efba74700000000
+------+---------+
| year | winner  |
+------+---------+
| 2018 | France  |
| 2014 | Germany |
| 2010 | Spain   |
| 2006 | Italy   |
+------+---------+
```

9. Log in using ssh to the host running HiveServer2 on the Compute cluster. You can find a host in the Cloudera Manager Admin Console by going to Clusters Compute 2 Hive Execution Service Instances .
10. Because this is a Kerberized environment, kinit the user:

```
[root@vpc_host-nnwznq-1 ~]# kinit hive
```

```
Password for hive@tut.myco.com:
```

11. Launch beeline:

- TLS is not enabled:

```
[root@vpc_host-nnwznq-1 ~]# beeline -u 'jdbc:hive2://localhost:10000/def
ault;principal=hive/vpc_host-nnwznq-1.tut.myco.com@tut.myco.com'
```

- Auto-TLS is enabled:

```
[root@vpc_host-nnwznq-1 ~]# beeline -u 'jdbc:hive2://localhost:10000/def
ault;ssl=true;sslTrustStore=<CMCA_TRUSTORE_LOC>.jks;trustStorePassword=c
loudera;principal=hive/vpc_host-nnwznq-1.tut.myco.com@tut.myco.com'
```

- Regular TLS is enabled:

```
[root@vpc_host-nnwznq-1 ~]# beeline -u 'jdbc:hive2://localhost:10000/def
ault;ssl=true;sslTrustStore=/etc/cdep-ssl-conf/CA_STANDARD/truststore.jk
s;trustStorePassword=cloudera;principal=hive/vpc_host-nnwznq-1.tut.myco.
com@tut.myco.com'
```

```
/CA_STANDARD/truststore.jks;trustStorePassword=cloudera;principal=hive/<
HiveServer2 Host URL>@VPC.CLOUDERA.COM'
```

12. Access the tables created through Impala in the previous section:

```
0: jdbc:hive2://localhost:10000/default> show databases;
```

```
INFO  : Compiling command(queryId=hive_20190309192547_09146fd4-58b9-4f60-
ad40-c9de3f98d470); Time taken: 0.987 seconds
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20190309192547_09146fd4-5
8b9-4f60-ad40-c9de3f98d470); Time taken: 0.041 seconds
INFO  : OK
+----------------+
| database_name  |
+----------------+
| default        |
| test_data      |
+----------------+
```

```
0: jdbc:hive2://localhost:10000/default> select * from test_table;
```

```
INFO  : Compiling command(queryId=hive_20190309192621_701914ad-0417-4639
-9209-335a63818b82): select * from test_table
command(queryId=hive_20190309192621_701914ad-0417-4639-9209-335a63818b82);
 Time taken: 0.38 seconds
+------------------+--------------------+
| test_table.year  | test_table.winner  |
+------------------+--------------------+
| 2018             | France             |
| 2014             | Germany            |
| 2010             | Spain              |
| 2006             | Italy              |
+------------------+--------------------+
```

13. Modify the table:

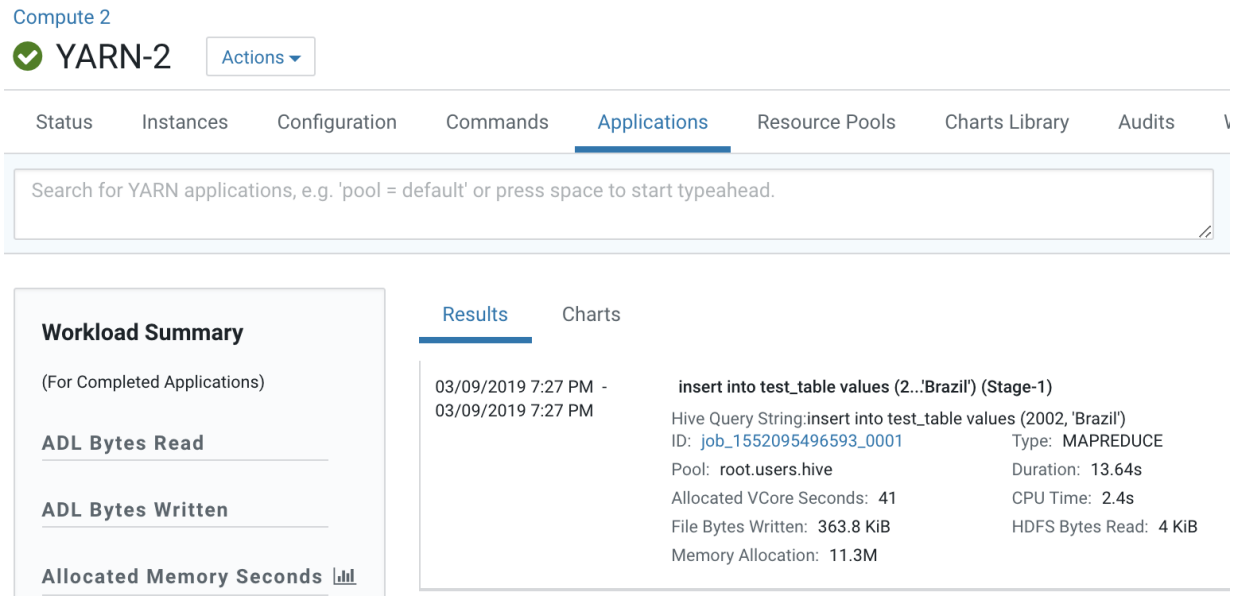
```
0: jdbc:hive2://localhost:10000/default> insert into test_table values (
2002, 'Brazil');
```

```
INFO  : Compiling
INFO  : Executing command(queryId=hive_20190309192705_218b79aa-aa94-4102-
95ab-a1d4bc7a0381): insert into test_table values (2002, 'Brazil')
WARN  :
INFO  : Query ID = hive_20190309192705_218b79aa-aa94-4102-95ab-a1d4bc7a0
381
INFO  : Total jobs = 3
INFO  : Launching Job 1 out of 3
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Submitting tokens for job: job_1552095496593_0001
INFO  : The url to track the job: https://vpc_host-nnwznq-1.tut.myco.com
:8090/proxy/application_1552095496593_0001/
INFO  : Starting Job = job_1552095496593_0001, Tracking URL = https://vp
c_host-nnwznq-1.tut.myco.com:8090/proxy/application_1552095496593_0001/
hdfs://ns1/user/hive/warehouse/test_data.db/test_table/.hive-staging_hiv
e_2019-03-09_19-27-05_193_3963732700280111926-1/-ext-10000 from hdfs://n
s1/user/hive/warehouse/test_data.db/test_table/.hive-staging_hive_2019-0
3-09_19-27-05_193_3963732700280111926-1/-ext-10002
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Loading data to table test_data.test_table from hdfs://ns1/user/h
ive/warehouse/test_data.db/test_table/.hive-staging_hive_2019-03-09_19-2
7-05_193_3963732700280111926-1/-ext-10000
INFO  : MapReduce Jobs Launched:
```

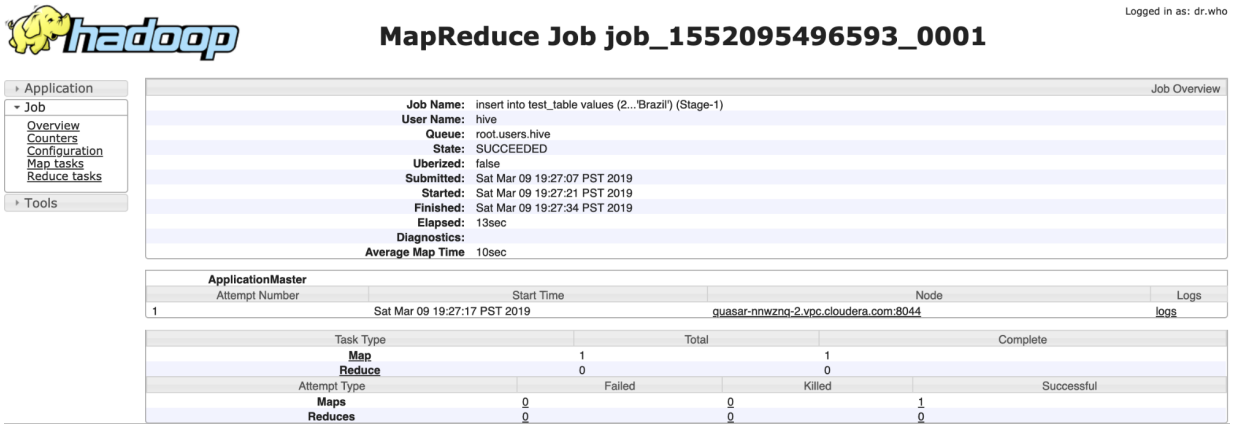
```
INFO  : Stage-Stage-1: Map: 1   Cumulative CPU: 2.4 sec   HDFS Read: 4113
 HDFS Write: 88 HDFS EC Read: 0 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 2 seconds 400 msec
INFO  : Completed executing command(queryId=hive_20190309192705_218b79aa-
aa94-4102-95ab-a1d4bc7a0381); Time taken: 31.097 seconds
INFO  : OK
1 row affected (31.434 seconds)
```

14. Verify and track the Yarn job submitted by the Hive Execution Service using the Cloudera Manager Admin Console by going to Clusters Compute 2 YARN 2 Applications .

YARN Job:



Map Reduce job:



15. Verify that new data was added to the table:

```
0: jdbc:hive2://localhost:10000/default> select * from test_table;
```

```
INFO  : Completed executing command(queryId=hive_20190309192752_38d2b5a0-
bf12-46b2-9a55-62e06beb9fb9); Time taken: 0.0 seconds
+------------------+--------------------+
| test_table.year  | test_table.winner  |
+------------------+--------------------+
| 2002             | Brazil             |
| 2018             | France             |
| 2014             | Germany            |
```

```
| 2010              | Spain              |
| 2006              | Italy              |
+-------------------+--------------------+
```

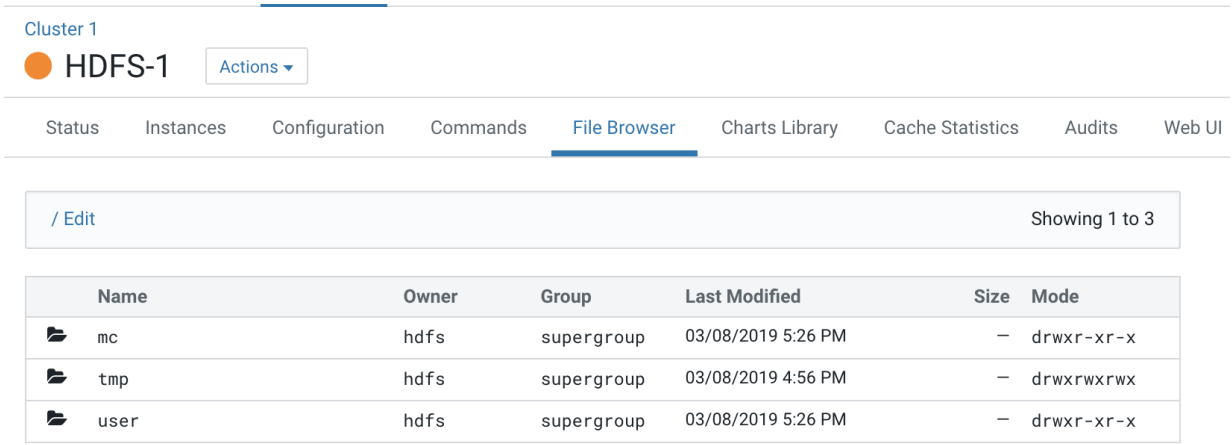
View HDFS directory structure of Compute clusters

Learn how to view the HDFS structure in a Base cluster from a Compute cluster.

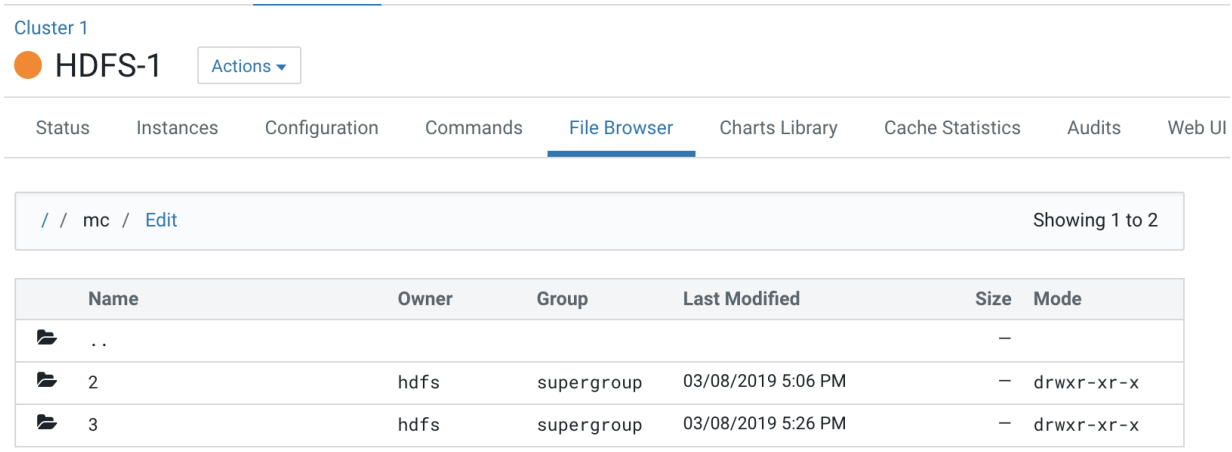
In a Virtual Private Cluster environment, some Compute clusters have a local HDFS. But, crucial directories (JHS for spark, logging for yarn) for compute services are created in the Base cluster HDFS.

One of the design assumptions of Compute clusters is that they would be transient and so the user should still have a way to access important logs after the Compute clusters have been destroyed.

1. Open the Cloudera Manager Admin Console and view the HDFS hierarchy on the Base cluster HDFS service by opening the File Browser: Cluster 1 HDFS-1 File Browser .



All the logs pertaining to Compute clusters are under the “mc” directory.



This Base cluster has 2 Compute clusters associated with it, Compute 1 and Compute 2.

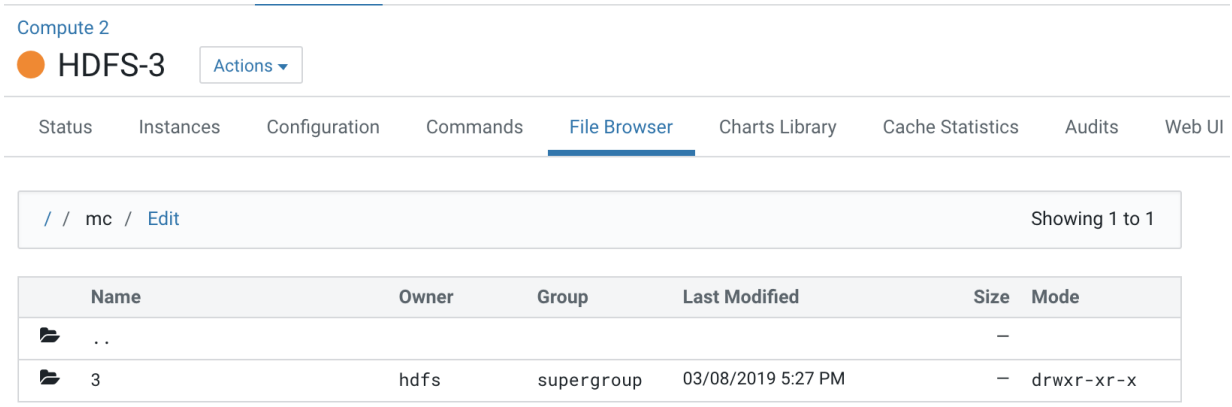
Each Compute cluster (based on its ID) gets a folder under this directory, so folder 2 belongs to Compute 1 and 3 belongs to Compute 2. The ID of the cluster can be identified from the URL used to access the cluster. Click on Compute 1 in the CM Cluster view and inspect the URL.

http://quasar-wfrgnj-1.vpc.cloudera.com:7180/cmf/clusters/2/status

The ID is the segment following /clusters in the URL. This is also the subfolder name under the /mc folder.

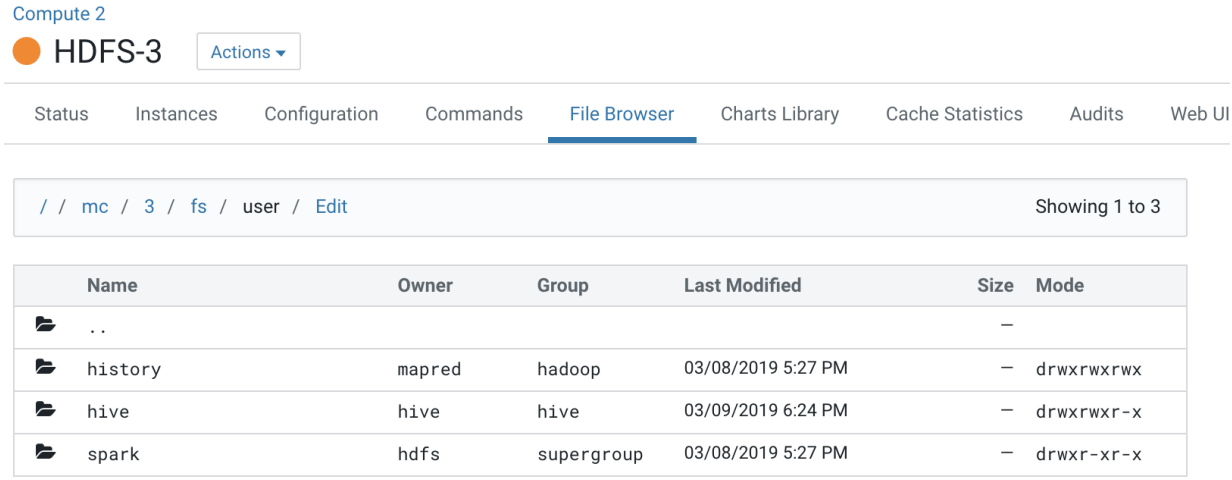
This is the directory where all the logs for services in Compute 1 are stored.

2. Navigate to the file browser of a Compute cluster.



Note that folder 2 which is dedicated for the Compute 1 cluster is not visible to the Compute 2 cluster.

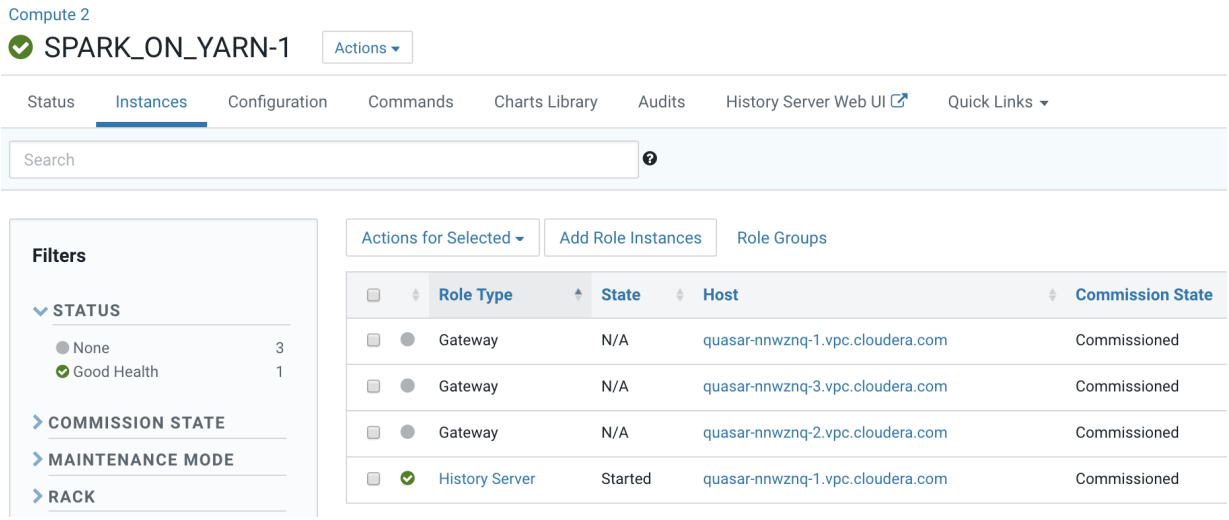
Navigating to folders below this hierarchy, you can see the folders created for services present on the Compute 2 cluster.



Insert data in test_table through Spark

Learn how to insert data using Spark.

1. Identify a host to start a spark-shell on the Compute cluster, Compute 2. Open the Cloudera Manager Admin Console and go to Clusters Compute 2 SPARK_ON_YARN-1 Instances .



2. Open a terminal session host <HiveServer2 Host URL>
3. Verify access for the user:

   Assign the user starting spark-shell to a Linux group that has create/insert access configured in Sentry. .e. hive.

```
[root@vpc_host-nnwznq-1 ~]# usermod -aG hive systest
```

   The user will also need to be created and added to the group on all the hosts of the Base cluster.
4. kinit the user:

```
[root@vpc_host-nnwznq-1 ~]# kinit -kt /cdep/keytabs/systest.keytab systest
```

5. Start spark-shell:

```
[root@vpc_host-nnwznq-1 ~]# sudo -u systest spark-shell
```

```
Setting default log level to "WARN".
Spark context Web UI available at http://vpc_host-nnwznq-1.tut.myco.com
:4043
Spark context available as 'sc' (master = yarn, app id = application_15520
95496593_0007).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.4.0-cdh6.2.0
      /_/
scala>
```

6. Insert data into the table:

```
scala> val insertData = sqlContext.sql("insert into test_data.test_table
 values (1998, 'France')")
```

```
19/03/10 00:21:26 WARN shims.HdfsUtils: Unable to inherit permissions fo
r file hdfs://ns1/user/hive/warehouse/test_data.db/test_table/part-00000-
f0fc5e0d-daa3-4fe2-9ded-d62e1ef45ca9-c000 from file hdfs://ns1/user/hive/
warehouse/test_data.db/test_table
```

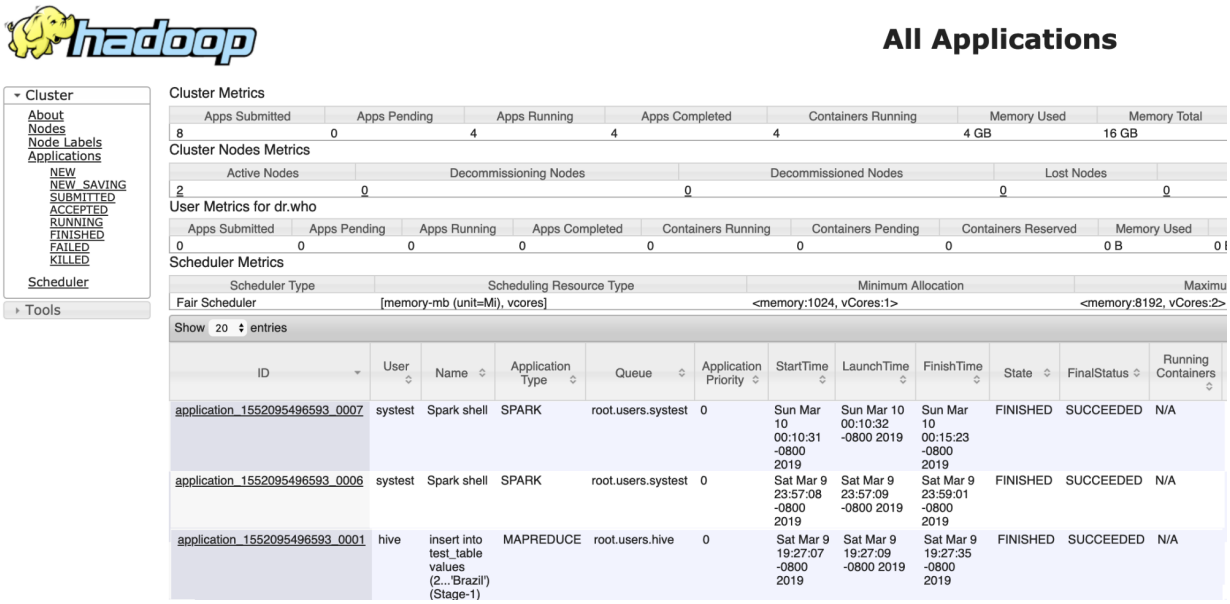
```
insertData: org.apache.spark.sql.DataFrame = []
```

```
scala> val tableTestData = sqlContext.sql("select * from test_data.test_
table")
tableTestData: org.apache.spark.sql.DataFrame = [year: int, winner: stri
ng]

scala> tableTestData.show()
+----+-------+

|year| winner|
+----+-------+
|2002| Brazil|
|2018| France|
|2014|Germany|
|2010|  Spain|
|2006|  Italy|
|1998| France|
+----+-------+
```

7. Verify and track the queries in the Yarn service application on the Compute cluster:



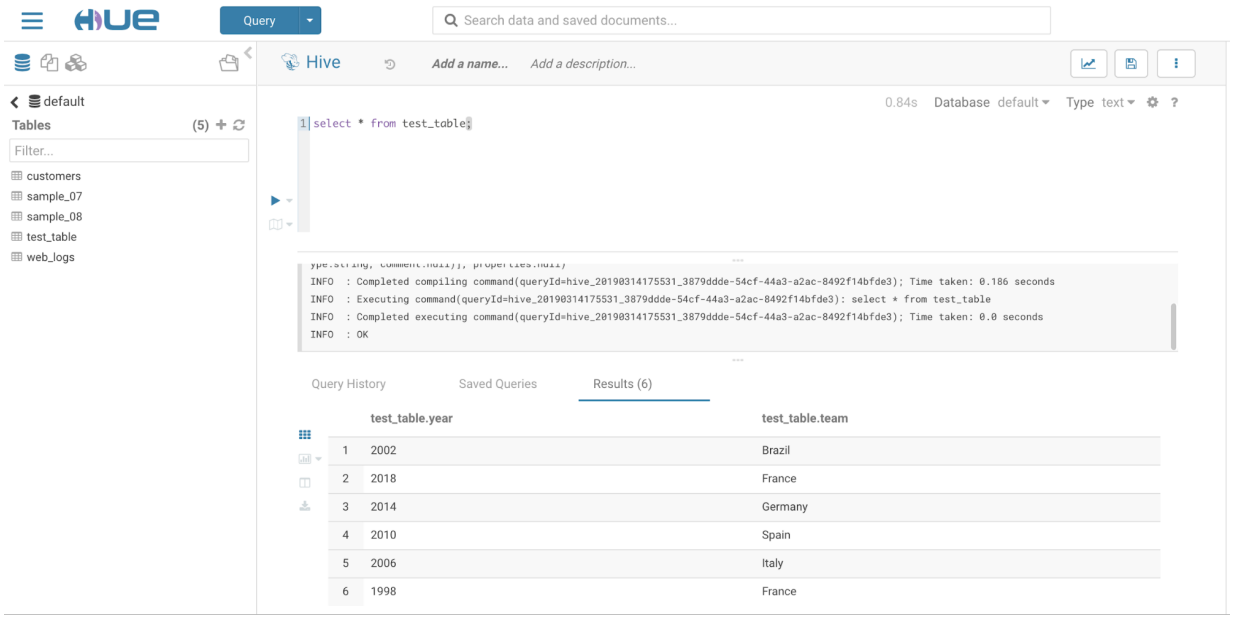
Hue in a Virtual Private Cluster Environment

Learn how to use Hue to view data in a Base cluster.

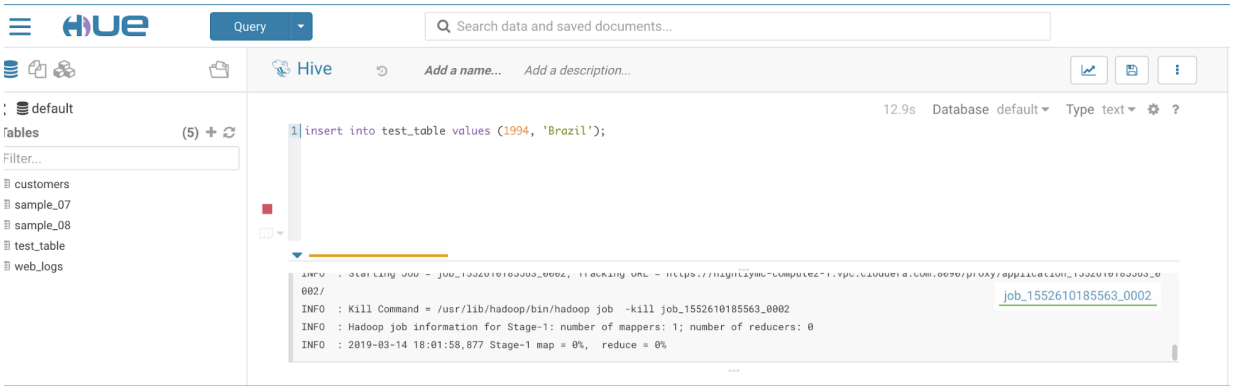
1. Login into Hue. Open the Cloudera Manager Admin Console and go to Clusters Compute Clusters Hue Hue Web UI Load Balanced Web UI .

   Note: If the Base cluster uses Sentry, ensure that the logged-in user is a part of a Linux group that has the necessary grants to access the Hive warehouse on the Base cluster.
2. Select Editor Hive .

   The tables display in the left panel.
3. Run a select query to display data in the test_table previously created:

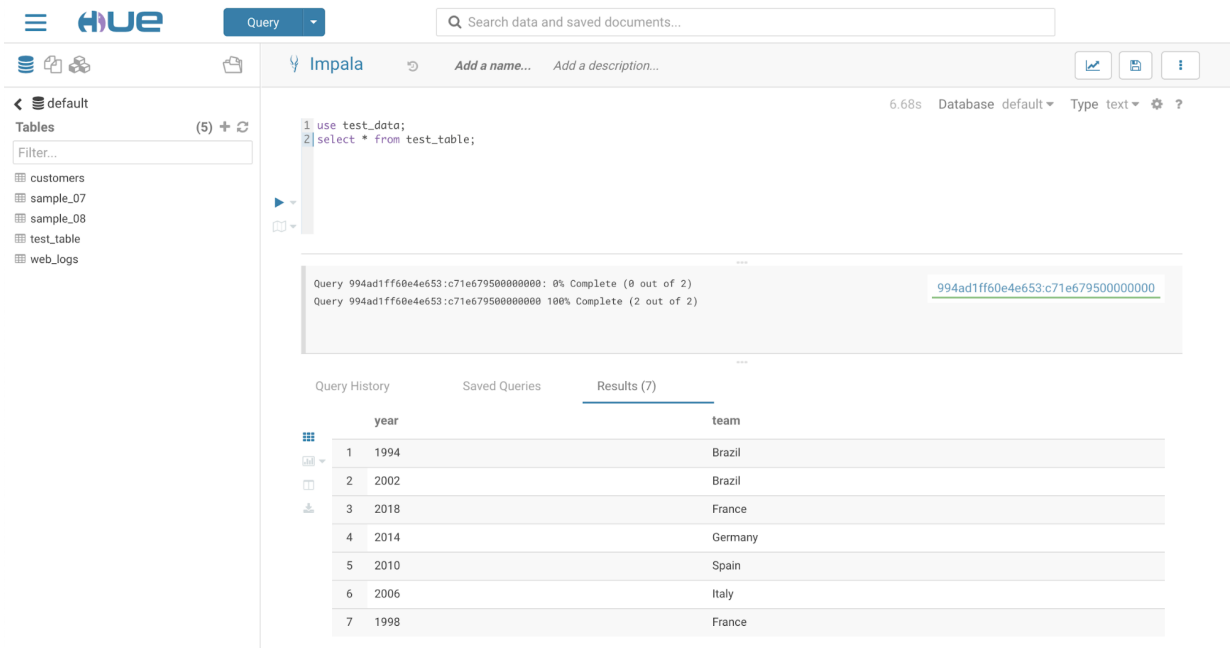


4. Insert data into the table:



5. Change the editor to Impala:

   Go to Hue Query Impala .
6. Run INVALIDATE METADATA query in the Impala editor.
7. Verify the data added from the Hive editor in the test_table shows up in the Impala editor:



Starting, Stopping, Refreshing, and Restarting a Cluster

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

Complete the steps below to start, stop, refresh, and restart a cluster.

Starting a Cluster

1. On the HomeStatus tab, click  to the right of the cluster name and select Start.
2. 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.



Note: The cluster-level Start action starts only Cloudera Runtime and other product services (Impala, Cloudera Search). It does not start the Cloudera Management Service. You must start the Cloudera Management Service separately if it is not already running.

Stopping a Cluster

1. On the HomeStatus tab, click  to the right of the cluster name and select Stop.
2. 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.

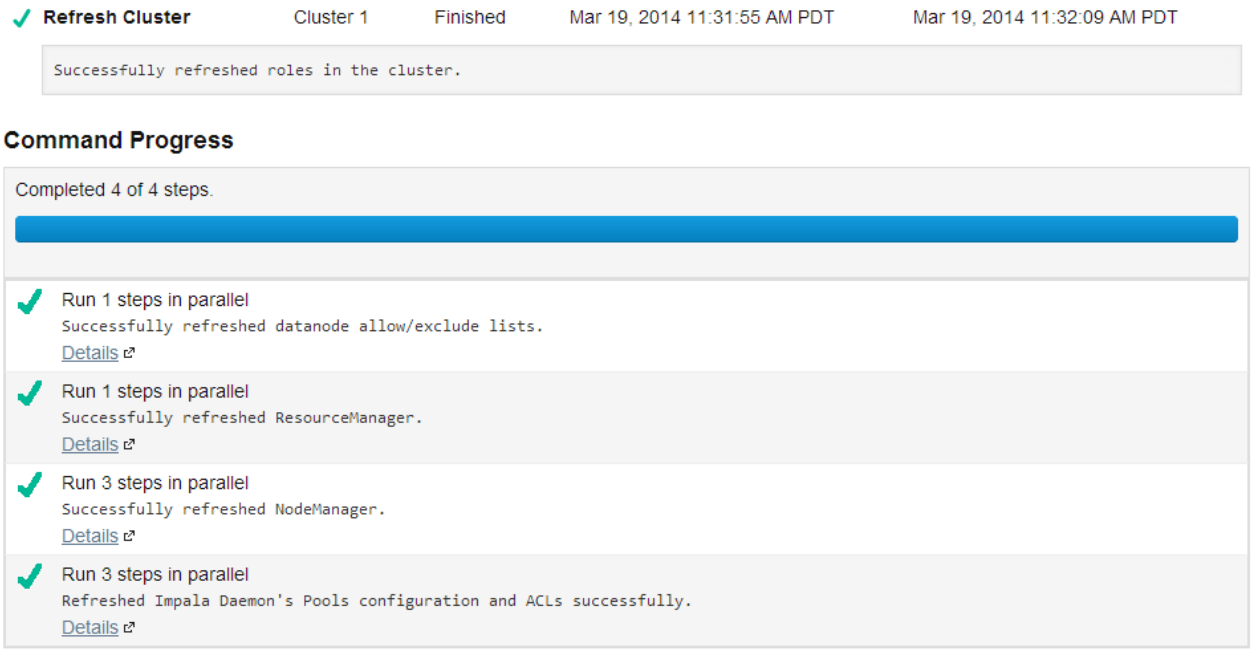


Note: The cluster-level Stop action does not stop the Cloudera Management Service. You must stop the Cloudera Management Service separately.

Refreshing a Cluster

Runs a cluster refresh action to bring the configuration up to date without restarting all services. For example, certain masters (for example NameNode and ResourceManager) have some configuration files (for example, fair-scheduler.xml, mapred_hosts_allow.txt, topology.map) that can be refreshed. If anything changes in those files then a refresh can be used to update them in the master.

Here is a summary of the operations performed in a refresh action:



To refresh a cluster, in the HomeStatus tab, click  to the right of the cluster name and select Refresh Cluster.

Restarting a Cluster

1. On the HomeStatus tab, click  to the right of the cluster name and select Restart.
2. Click Restart that appears in the next screen to confirm. If you have enabled high availability for HDFS, you can choose Rolling Restart instead to minimize cluster downtime. The Command Details window shows the progress of stopping services.

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

Renaming a Cluster

Renaming the cluster allows you to change the display name of the cluster after you have created the cluster. However, this does not change the cluster's actual, internal name.

About this task

Cloudera Manager does not allow you to change the internal name of the cluster because the internal name is used as part of different identifiers throughout the cluster, and it is not possible to change all the configurations safely.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. Log into Cloudera Manager as an Administrator.
2. On the Cloudera Manager home page, navigate to the Status tab, click ⋮ to the right of the cluster name and select Rename Cluster from the list view.
3. Enter the new cluster name and click Rename Cluster.



Note:

In the Configuration tab, the cluster name still appears as the original cluster name that was used when the cluster was created.

Managing Hosts

How to use Cloudera Manager to configure and manage the hosts in your clusters.

Viewing Host Status

You can view summary information about the hosts managed by Cloudera Manager. You can view information for all hosts, the hosts in a cluster, or individual hosts.

Viewing All Hosts

To display summary information about all the hosts managed by Cloudera Manager, click HostsAll Hosts in the left menu. The **All Hosts** page displays with a list of all the hosts managed by Cloudera Manager.

The list of hosts shows the overall status of the Cloudera Manager-managed hosts in your cluster.

- The information provided varies depending on which columns are selected. To change the columns, click the Columns: *N* Selected drop-down and select the checkboxes next to the columns to display.
- Click › to the left of the number of roles to list all the role instances running on that host.
- Filter the hosts list by entering search terms (hostname, IP address, or role) in the search box separated by commas or spaces. Use quotes for exact matches (for example, strings that contain spaces, such as a role name) and brackets to search for ranges. Hosts that match any of the search terms are displayed. For example:

```
hostname[1-3], hostname8 hostname9, "hostname.example.com"
hostname.example.com “HDFS DataNode”
```

- You can also search for hosts by selecting a value from the facets in the Filters section at the left of the page. Click the Filters toggle to show or hide the Filters section.
- If the agent heartbeat and health status properties are configured as follows:
  - Send Agent heartbeat every *X*
  - Set health status to Concerning if the Agent heartbeats fail *Y*
  - Set health status to Bad if the Agent heartbeats fail *Z*

  The value *V* for a host's Last Heartbeat facet is computed as follows:

  - *V* < *X* * *Y* = Good
  - *V* >= *X* * *Y* and <= *X* * *Z* = Concerning
  - *V* >= *X* * *Z* = Bad

Viewing the Hosts in a Cluster

Do one of the following:

- Select Clusters *CLUSTER NAME* Hosts .

- In the Home screen, click Hosts in a full form cluster table.

The **All Hosts** page displays with a list of the hosts filtered by the cluster name.

Viewing Individual Hosts

You can view detailed information about an individual host—resources (CPU/memory/storage) used and available, which processes it is running, details about the host agent, and much more—by clicking a host link on the **All Hosts** page.

Related Information

Configuring Cloudera Manager Agents

Adding a Host to a Cluster

Steps to add hosts to a cluster.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

You can add one or more hosts to your cluster using the Add Hosts wizard, which installs the Oracle JDK, Cloudera Runtime, and Cloudera Manager Agent software. After the software is installed and the Cloudera Manager Agent is started, the Agent connects to the Cloudera Manager Server and you can use the Cloudera Manager Admin Console to manage and monitor Cloudera Runtime on the new host.

The Add Hosts wizard does not create roles on the new host; once you have successfully added the host(s) you can either add roles, one service at a time, or apply a host template, which can define role configurations for multiple roles.

Important:

- The hostname must be in lowercase, for example, *HOST01.EXAMPLE.COM*. If you use uppercase letters in any hostname, the cluster services will not start after enabling Kerberos.
- Unqualified hostnames (short names) must be unique in a Cloudera Manager instance. For example, you cannot have both *HOST01.EXAMPLE.COM* and *HOST01.STANDBY.EXAMPLE.COM* managed by the same Cloudera Manager Server.
- All hosts in a single cluster must be running the same version of Cloudera Runtime.
- When you add a new host, you must install the same version of Cloudera Runtime to enable the new host to work with the other hosts in the cluster. The installation wizard lets you select the version of Cloudera Runtime to install, and you can choose a custom repository to ensure that the version you install matches the version on the other hosts.
- If you are managing multiple clusters, select the version of Cloudera Runtime that matches the version in use on the cluster where you plan to add the new host.
- When you add a new host, the following occurs:
  - YARN topology.map is updated to include the new host
  - Any service that includes topology.map in its configuration—Flume, Hive, Hue, Oozie, Solr, Spark, YARN—is marked stale

  At a convenient point after adding the host you should restart the stale services to pick up the new configuration.
- Ensure that you configure your new nodes to adhere strictly to the Operating System Requirements for your current Runtime cluster version.

Use one of the following methods to add a new host:

* Azul OpenJDK, OpenJDK 8, OpenJDK 11, and OpenJDK 17 are TCK certified for CDP.

Using the Add Hosts Wizard to Add Hosts

You can use the Add Hosts wizard to install Cloudera Runtime, Impala, and the Cloudera Manager Agent on a host. There are additional considerations when TLS is enabled.

Important: If you have enabled TLS in the cluster you must use one of the following options:

- If you enabled TLS using Auto TLS, then you can use the Add Hosts wizard in Cloudera Manager.
- If you have manually enabled TLS, you can either:
  - Install the Cloudera Manager Agent on the new hosts manually. See Alternate Method of Installing Cloudera Manager Agent without Disabling TLS on page 36, below.
  - Disable TLS before starting the Add Hosts wizard. You can then re-enable TLS after adding the hosts. See Disable TLS Encryption or Authentication on page 36, below.

Disable TLS Encryption or Authentication

Before you begin

Warning: If you previously enabled TLS using Auto-TLS through the Cloudera Manager Web UI, you must comment out the auto-generated cm_init.txt block in /etc /default/cloudera-scm-server before you restart the Cloudera Manager Server. If you skip this step, Cloudera Manager automatically re-enables TLS settings upon restart.

If you have enabled TLS encryption or authentication for the Cloudera Manager Agents, you must disable both of them before starting the Add Hosts wizard. Otherwise, skip to the next step.

If you perform this step, then skip step 2 (Alternate Method of Installing Cloudera Manager Agent without Disabling TLS). If you skip step 1 and perform step 2, then continue to step 3 (Using the Add Hosts Wizard to Add Hosts).

Important: This step temporarily puts the existing cluster hosts in an unmanageable state; they are still configured to use TLS and so cannot communicate with the Cloudera Manager Server. Roles on these hosts continue to operate normally, but Cloudera Manager is unable to detect errors and issues in the cluster and reports all hosts as being in bad health. To work around this issue, you can manually install the Cloudera Manager Agent on the new host. See Alternate Method of Installing Cloudera Manager Agent without Disabling TLS on page 36.

1. From the Administration tab, select Settings.
2. Select the Security category.
3. Disable TLS by clearing the following options:
   - Use TLS Encryption for Agents
   - Use TLS Authentication of Agents to Server
4. Click Save Changes to save the settings.
5. Log in to the Cloudera Manager Server host through SSH.
6. Check for the presence of the cm_init.txt auto-initialization script in the Cloudera Manager Server default environment file:

   ```
   grep cm_init.txt /etc/default/cloudera-scm-server
   ```

   If the command returns output referencing cm_init.txt, open /etc/default/cloudera-scm-server in a text editor and comment out (add # to the beginning of each line) the entire auto-generated text block containing cm_init.txt.
7. Restart the Cloudera Manager Server:

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

Alternate Method of Installing Cloudera Manager Agent without Disabling TLS

If you have TLS encryption or authentication enabled in your cluster, you must either disable TLS during the installation, or install the Cloudera Manager Agent manually using the following procedure:

1. Copy the repository configuration file from an existing host in the cluster to the new host. For example:

OS	Command
RHEL	`sudo scp mynode.example.com:/etc/yum.repos.d/cloudera-manager.repo /etc/yum.repos.d/cloudera-manager.repo`
SLES	`sudo scp mynode.example.com:/etc/zypp/zypper.conf/cloudera-cm.repo /etc/zypp/zypper.conf/cloudera-cm.repo`
Ubuntu or Debian	`sudo scp mynode.example.com:/etc/apt/sources.list.d/cloudera.list /etc/apt/sources.list.d/cloudera.list`

2. Remove cached package lists and other transient data by running the following command:

OS	Command
RHEL	`sudo yum clean all`
SLES	`sudo zypper clean --all`
Ubuntu or Debian	`sudo apt-get clean`

3. Install the JDK package from the Cloudera Manager repository. Install the same version as is used on other cluster hosts. Only JDK 1.8 is supported:

Table 1: Oracle JDK 1.8

OS	Command
RHEL	`sudo yum install jdk1.8.0_144-cloudera`
SLES	`sudo zypper install jdk1.8.0_144-cloudera`
Ubuntu or Debian	`sudo apt-get install jdk1.8.0_144-cloudera`

Open JDK

RHEL

OpenJDK 8*

```
sudo yum install java-1.8.0-openjdk-devel
```

OpenJDK 11*

```
sudo yum install java-11-openjdk-devel
```

OpenJDK 17*

```
sudo yum install java-17-openjdk-devel
```

Ubuntu

OpenJDK 8*

```
sudo apt-get install openjdk-8-jdk
```

OpenJDK 11*

```
sudo apt install openjdk-11-jdk
```

OpenJDK 17*

```
sudo apt install openjdk-17-jdk
```

SLES

OpenJDK 8*

```
sudo zypper install java-1_8_0-openjdk-devel
```

OpenJDK 11*

```
zypper install java-11-openjdk-devel
```

OpenJDK 17*

```
sudo zypper --non-interactive install java-17-openjdk-devel
```

4. Set up the TLS certificates using the same procedure that was used to set them up on other cluster hosts. See Configuring TLS Encryption for Cloudera Manager Using Auto-TLS. If you have set up a custom truststore, copy that file from an existing host to the same location on the new host.
5. Install the Cloudera Manager Agent:

OS	Command
RHEL	`sudo yum install cloudera-manager-agent`
SLES	`sudo zypper install cloudera-manager-agent`
Ubuntu or Debian	`sudo apt-get install cloudera-manager-agent`

6. Copy the Cloudera Manager Agent configuration file from an existing cluster host that is already configured for TLS to the same location on the new host. For example:

```
sudo scp mynode.example.com:/etc/cloudera-scm-agent/config.ini /etc/clou
dera-scm-agent/config.ini
```

7. Create and secure the file containing the password used to protect the private key of the Agent:
   a. Use a text editor to create a file called agentkey.pw that contains the password. Save the file in the /etc/cloudera-scm-agent directory.
   b. Change ownership of the file to root:

```
sudo chown root:root /etc/cloudera-scm-agent/agentkey.pw
```

c. Change the permissions of the file:

```
sudo chmod 440 /etc/cloudera-scm-agent/agentkey.pw
```

8. Start the Agent on the new host:

```
sudo service cloudera-scm-agent start
```

9. Log in to Cloudera Manager and go to HostsAll Hosts page and verify that the new host is recognized by Cloudera Manager.

Add Hosts to an Existing Cluster

1. Click the Hosts tab.
2. Click the Add Hosts button.
3. Select Add hosts to cluster.
4. If the cluster uses Kerberos authentication, ensure that the Kerberos packages are installed on the new hosts. If necessary, use the package commands provided on the Add Hosts screen to install these packages.
5. Select the cluster where you want to add the host from the drop-down list.
6. Click Continue.

   The Specify Hosts page displays. You can either add a new host to the cluster, or add an existing managed host to the cluster.

   Do one of the following:

   - Add a new host:

     a. On the Specify Hosts page, enter a host name or pattern (click "using patterns" for more information) to search for new hosts to add to the cluster.

        A list of matching hosts displays.

     b. Select the hosts that you want to add.

     c. Click Continue.

     d. Select the Repository Location where Cloudera Manager can find the software to install on the new hosts. Select Public Cloudera Repository or Custom Repository and enter the URL of a custom repository available on your local network.

     e. Click Continue.

     f. Follow the instructions in the wizard to install the Oracle JDK.

     g. Enter Login Credentials:

        1. Select root for the root account, or select Another user and enter the username for an account that has password-less sudo privileges.
        2. Select an authentication method:

           - If you choose password authentication, enter and confirm the password.
           - If you choose public-key authentication, provide a passphrase and path to the required key files.

           You can modify the default SSH port if necessary.
        3. Specify the maximum number of host installations to run at once. The default and recommended value is 10. You can adjust this based on your network capacity.
        4. Click Continue.

           The Install Agents page displays and Cloudera Manager installs the Agent software on the new hosts.
        5. When the agent installation finishes, click Continue.

   - Add an existing managed host:

     a. Click the Currently Managed Hosts tab.

        A list of hosts previously added to Cloudera Manager displays.

   b. Select the hosts that you want to add to the cluster.
   c. Click Continue.
7. Cloudera Manager begins to install the Cloudera Runtime parcels.
8. When the parcel installation finishes, click Continue.
9. The Host Inspector runs and displays any problems with the hosts. Correct the problems before continuing.
10. After correcting any problems, click Continue.
11. To add role instances to the hosts now or select None to add them later.
12. To add roles now:
    a. Select an existing host template, or create a new one.
    b. To create a new host template, click the Create... button. The Create New Host Template screen opens.. See Host Templates on page 43 for details on how you select the role groups that define the roles that should run on a host. After you have created the template, it will appear in the list of host templates from which you can choose.
    c. Select the host template you want to use.
    d. By default Cloudera Manager will automatically start the roles specified in the host template on your newly added hosts. To prevent this, uncheck the option to start the newly-created roles.
13. When the wizard is finished, you can verify the Agent is connecting properly with the Cloudera Manager Server by clicking the Hosts tab and checking the health status for the new host. If the Health Status is Good and the value for the Last Heartbeat is recent, then the Agent is connecting properly with the Cloudera Manager Server.

If you did not specify a host template during the Add Hosts wizard, then no roles will be present on your new hosts until you add them. You can do this by adding individual roles under the Instances tab for a specific service, or by using a host template. See Adding a Role Instance for information about adding roles for a specific service. See Host Templates to create a host template that specifies a set of roles (from different services) that should run on a host.

Add New Hosts To Cloudera Manager

This option allows you to add a host, but without adding them to a specific cluster. Later, you can use these hosts to create new clusters or expand existing clusters.

1. Click the Hosts tab.
2. Click the Add Hosts button.
3. Select Add hosts to Cloudera Manager
4. Specify the hosts to add:
    a. On the Specify Hosts page, enter a host name or pattern (click "using patterns" for more information) to search for new hosts to add to the cluster.

       A list of matching hosts displays.
    b. Select the hosts that you want to add.
    c. Click Continue.
    d. Select the Repository Location where Cloudera Manager can find the software to install on the new hosts. Select Public Cloudera Repository or Custom Repository and enter the URL of a custom repository available on your local network.
    e. Click Continue.
    f. Follow the instructions in the wizard to install the Oracle JDK.
    g. Enter Login Credentials:
        1. Select root for the root account, or select Another user and enter the username for an account that has password-less sudo privileges.
        2. Select an authentication method:
            - If you choose password authentication, enter and confirm the password.
            - If you choose public-key authentication, provide a passphrase and path to the required key files.

           You can modify the default SSH port if necessary.

3. Specify the maximum number of host installations to run at once. The default and recommended value is 10. You can adjust this based on your network capacity.
4. Click Continue.

   The Install Agents page displays and Cloudera Manager installs the Agent software on the new hosts.
5. When the agent installation finishes, click Continue.

5. If the cluster uses Kerberos authentication, ensure that the Kerberos packages are installed on the new hosts. If necessary, use the package commands provided on the Add Hosts screen to install these packages.
6. Select the Repository Location where Cloudera Manager can find the software to install on the new hosts. Select Public Cloudera Repository or Custom Repository and enter the URL of a custom repository available on your local network.
7. Follow the instructions in the wizard to install the Oracle JDK.
8. Enter Login Credentials:

   a. Select root for the root account, or select Another user and enter the username for an account that has password-less sudo privileges.
   b. Select an authentication method:

      - If you choose password authentication, enter and confirm the password.
      - If you choose public-key authentication, provide a passphrase and path to the required key files.

      You can modify the default SSH port if necessary.
   c. Specify the maximum number of host installations to run at once. The default and recommended value is 10. You can adjust this based on your network capacity.
   d. Click Continue.

      The Install Agents page displays and Cloudera Manager installs the Agent software on the new hosts.
   e. When the agent installation finishes, click Continue.
9. The Host Inspector runs and displays any problems with the hosts. Correct the problems before continuing.
10. After correcting any problems, click Continue.

Enable TLS Encryption or Authentication

If you previously enabled TLS security on your cluster, you must re-enable the TLS options on the Administration page and also configure TLS on each new host after using the Add Hosts wizard. Otherwise, you can ignore this step. For instructions, see Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS.

Enable TLS/SSL for cluster Components

If you have previously enabled TLS/SSL on your cluster, and you plan to start these roles on this new host, make sure you install a new host certificate to be configured from the same path and naming convention as the rest of your hosts. Since the new host and the roles configured on it are inheriting their configuration from the previous host, ensure that the keystore or truststore passwords and locations are the same on the new host. For instructions on configuring TLS/SSL, see Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS.

Enable Kerberos

If you have previously enabled Kerberos on your cluster:

1. Install the packages required to kinit on the new host (see the list in Enabling Kerberos Authentication for CDH).
2. If you have set up Cloudera Manager to manage krb5.conf, it will automatically deploy the file on the new host. Note that Cloudera Manager will deploy krb5.conf only if you use the Kerberos wizard. If you have used the API, you will need to manually perform the commands that the wizard calls.

   If Cloudera Manager does not manage krb5.conf, you must manually update the file at /etc/krb5.conf.

Adding a Host by Installing the Packages Using Your Own Method

If you used a different mechanism to install the JDK, Cloudera Runtime, and Cloudera Manager Agent packages, you can use that same mechanism to install the JDK, Cloudera Runtime, Cloudera Manager Agent packages and then start the Cloudera Manager Agent.

1. Install the Oracle JDK, Cloudera Runtime, and Cloudera Manager Agent packages using your own method. For instructions on installing these packages, see Step 3: Install Cloudera Manager Server.
2. After installation is complete, start the Cloudera Manager Agent. For instructions, see Starting, Stopping, and Restarting Cloudera Manager Agents on page 98.
3. After the Agent is started, you can verify the Agent is connecting properly with the Cloudera Manager Server by clicking the Hosts tab and checking the health status for the new host. If the Health Status is Good and the value for the Last Heartbeat is recent, then the Agent is connecting properly with the Cloudera Manager Server.
4. If you have enabled TLS security on your cluster, you must enable and configure TLS on each new host. Otherwise, ignore this step.
   a. Enable and configure TLS on each new host by specifying 1 for the use_tls property in the /etc/cloudera-scm-agent/config.ini configuration file.
   b. Configure TLS security on the new hosts by following the instructions in Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS.
5. If you have previously enabled TLS/SSL on your cluster, and you plan to start these roles on this new host, make sure you install a new host certificate to be configured from the same path and naming convention as the rest of your hosts. Since the new host and the roles configured on it are inheriting their configuration from the previous host, ensure that the keystore or truststore passwords and locations are the same on the new host. For instructions on configuring TLS/SSL, see Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS.
6. If you have previously enabled Kerberos on your cluster:
   a. Install the packages required to kinit on the new host (see the list in Enabling Kerberos Authentication for CDH).
   b. If you have set up Cloudera Manager to manage krb5.conf, it will automatically deploy the file on the new host. Note that Cloudera Manager will deploy krb5.conf only if you use the Kerberos wizard. If you have used the API, you will need to manually perform the commands that the wizard calls.

      If Cloudera Manager does not manage krb5.conf, you must manually update the file at /etc/krb5.conf.

Parcels

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

In the Parcels tab you can download, distribute, and activate available parcels to your cluster. You can use parcels to add new products to your cluster, or to upgrade products you already have installed.

Configuring Hosts

The Configuration tab lets you set properties related to parcels and to resource management, and also monitoring properties for the hosts under management.

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

The configuration settings you make here will affect all your managed hosts. You can also configure properties for individual hosts by clicking on the host in the **All Hosts** page, which will override the global properties set here).

To edit the default configuration properties for hosts, click the Configuration tab.

Related Information

Modifying Configuration Properties Using Cloudera Manager

Viewing Host Role Assignments

You can view the assignment of roles to hosts as follows:

1. In the left menu, click HostsRoles.
2. Click a cluster name or All Clusters.

Host Templates

The **Host Templates** page lets you create and manage host templates, which provide a way to specify a set of role configurations that should be applied to a host.

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

Host templates let you designate a set of role groups that can be applied in a single operation to a host or a set of hosts. This significantly simplifies the process of configuring new hosts when you need to expand your cluster.



Important: A host template can only be applied on a host with a version of Cloudera Runtime that matches the Cloudera Runtime version running on the cluster to which the host template belongs.

You can create and manage host templates by clicking HostsHost Templates.

Templates are not required; Cloudera Manager assigns roles and role groups to the hosts of your cluster when you perform the initial cluster installation. However, if you want to add new hosts to your cluster, a host template can make this much easier.

If there are existing host templates, they are listed on the page, along with links to each role group included in the template.

If you are managing multiple clusters, you must create separate host templates for each cluster, as the templates specify role configurations specific to the roles in a single cluster. Existing host templates are listed under the cluster to which they apply.

- You can click a role group name to be taken to the Edit configuration page for that role group, where you can modify the role group settings.
- From the Actions menu associated with the template you can edit the template, clone it, or delete it.

Creating a Host Template

When you create a host template, you choose a name for the template and select appropriate role groups for each role.

About this task

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

Procedure

1. Click HostsHost Templates.
2. From the **Host Templates** page, click Create.
   The **Create New Host Template** pop-up window appears.
3. Type a name for the template.
4. For each role, select the appropriate role group. There may be multiple role groups for a given role type — you want to select the one with the configuration that meets your needs.
5. Click Create to create the host template.

Editing a Host Template

You can edit the name of a host template, in addition to any of the role group selections.

About this task

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

Procedure

1. Click HostsHost Templates.
2. Pull down the Actions menu for the template you want to modify, and click Edit.
   The **Edit Host Template** window appears. This page is identical to the Create New Host Template page. You can modify the template name or any of the role group selections.
3. Click OK when you have finished.

Deleting a Host Template

You can delete a Host Template using the Cloudera Manager Admin Console.

Procedure

1. Open the Cloudera Manager Admin Console.
2. Click HostsHost Templates
3. In the row containing the Host Template that you want to delete, click the actions menu on the right and select Delete.

Applying a Host Template to a Host

You can use a host template to apply configurations for multiple roles in a single operation.

About this task

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

You can apply a template to a host that has no roles on it, or that has roles from the same services as those included in the host template. New roles specified in the template that do not already exist on the host will be added. A role on the host that is already a member of the role group specified in the template will be left unchanged. If a role on the host matches a role in the template, but is a member of a different role group, it will be moved to the role group specified by the template.

For example, suppose you have two role groups for a DataNode (DataNode Default Group and DataNode (1)). The host has a DataNode role that belongs to DataNode Default Group. If you apply a host template that specifies the DataNode (1) group, the role on the host will be moved from DataNode Default Group to DataNode (1).

However, if you have two instances of a service, such as MapReduce (for example, mr1 and mr2) and the host has a TaskTracker role from service mr2, you cannot apply a TaskTracker role from service mr1.

A host may have no roles on it if you have just added the host to your cluster, or if you decommissioned a managed host and removed its existing roles.

Also, the host must have the same version of CDH installed as is running on the cluster whose host templates you are applying.

If a host belongs to a different cluster than the one for which you created the host template, you can apply the host template if the "foreign" host either has no roles on it, or has only management roles on it. When you apply the host template, the host will then become a member of the cluster whose host template you applied. The following instructions assume you have already created the appropriate host template.

Procedure

1. Click HostsAll Hosts.
2. Select the host(s) to which you want to apply your host template.
3. From the Actions for Selected menu, select Apply Host Template.

4. In the pop-up window that appears, select the host template you want to apply.
5. Optionally you can have Cloudera Manager start the roles created per the host template. To enable this, check the box.
6. Click Confirm to initiate the action.

Hosts Disks Overview

How to view the status of all disks in a cluster.

In the left menu, click HostsDisks Overview to display an overview of the status of all disks in the deployment. The statistics exposed match or build on those in iostat, and are shown in a series of histograms that by default cover every physical disk in the system.

Adjust the endpoints of the time line to see the statistics for different time periods. Specify a filter in the box to limit the displayed data. For example, to see the disks for a single rack rack1, set the filter to: logicalPartition = false and rackId = "rack1" and click Filter. Click a histogram to drill down and identify outliers. Mouse over the graph and click ⤢ to display additional information about the chart.

Deleting Hosts

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

You can remove a host from a cluster in two ways:

- Delete the host entirely from Cloudera Manager.
- Remove a host from a cluster, but leave it available to other clusters managed by Cloudera Manager.

Both methods decommission the hosts, delete roles, and remove managed service software, but preserve data directories.

Deleting a Host from Cloudera Manager

To delete a host from Cloudera Manager, first decommission the host and then remove it.

About this task

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. In the Cloudera Manager Admin Console, go to Hosts All Hosts.
2. Select the hosts to delete.
3. Select Actions for SelectedHosts Decommission.
4. Stop the Agent on the host.
5. In the Cloudera Manager Admin Console, go to Hosts All Hosts.
6. Reselect the hosts you selected in Step 2.
7. Select Actions for SelectedRemove from Cloudera Manager.

Removing a Host From a Cluster

Removing a host from a cluster leaves the host managed by Cloudera Manager and preserves the Cloudera Management Service roles (such as the Events Server, Host Monitor,and so on).

About this task

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. In the Cloudera Manager Admin Console, click the Hosts tab.
2. Select the hosts to delete.
3. Select Actions for SelectedRemove From Cluster. The **Remove Hosts From Cluster** dialog box displays.
4. Leave the selections to decommission roles and skip removing the Cloudera Management Service roles. Click Confirm to proceed with removing the selected hosts.

Stopping All the Roles on a Host

You can stop all of the roles on a host from the **Hosts** page.

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 ClustersHosts or HostsAll Hosts.
2. Select one or more hosts on which to stop all roles.
3. Select Actions for SelectedStop Roles on Hosts.

Starting All the Roles on a Host

You can start all the roles on a host from the **Hosts** page.

About this task

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

Procedure

1. Click the Hosts tab.
2. Select one or more hosts on which to start all roles.
3. Select Actions for SelectedStart Roles on Hosts.

Changing Hostnames

After you have installed Cloudera Manager and created a cluster, you may need to update the names of the hosts running the Cloudera Manager Server or cluster services.

About this task

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.



Important:

- The process described here requires Cloudera Manager and cluster downtime.
- If any user-created scripts reference specific hostnames, those must also be updated.
- Due to the length and complexity of the following procedure, changing cluster hostnames is not recommended by Cloudera.

To update a deployment with new hostnames, follow these steps:

Procedure

1. Verify if TLS/SSL certificates have been issued for any of the services and make sure to create new TLS/SSL certificates in advance for services protected by TLS/SSL.
2. Export the Cloudera Manager configuration using one of the following methods:
   - Open a browser and go to this URL http://*CM_HOSTNAME*:7180/api/*API_VERSION*/cm/deployment. Save the displayed configuration.
   - From terminal type:

     $ curl -u admin:admin http://*CM_HOSTNAME*:7180/api/*API_VERSION*/cm/deployment > cme-cm-export.json

     If Cloudera Manager SSL is in use, specify the -k switch and the port number as 7183:

     $ curl -k -u admin:admin https://*CM_HOSTNAME*:7183/api/*API_VERSION*/cm/deployment > cme-cm-export.json

     where *CM_HOSTNAME* is the name of the Cloudera Manager host and *API_VERSION* is the correct version of the API for the version of Cloudera Manager you are using. For example, http://tcdn5-1.ent.cloudera.com:7180/api/v57/cm/deployment.
3. Stop all services on the cluster.
4. Stop the Cloudera Management Service.
5. Stop the Cloudera Manager Server.
6. Stop the Cloudera Manager Agents on the hosts that you want to change the hostname of.
7. Back up the Cloudera Manager Server database using mysqldump, pg_dump, or another preferred backup utility. Store the backup in a safe location.
8. Update names and principals:
   a) Update the target hosts using standard per-OS/name service methods (/etc/hosts, dns, /etc/sysconfig/network, hostname, and so on). Ensure that you remove the old hostname.
   b) If you are changing the hostname of the host running Cloudera Manager Server do the following:
      1. Change the hostname per Step 8.a.
      2. Update the Cloudera Manager hostname in /etc/cloudera-scm-agent/config.ini on all Agents.
   c) If the cluster is configured for Kerberos security, do the following:
      1. Remove the old hostname cluster principals.
         - If you are using an MIT KDC, remove old hostname cluster service principals from the KDC database using one of the following:
           - Use the delprinc command within kadmin.local interactive shell.

             OR
           - From the command line:

             ```
             kadmin.local -q "listprincs" | grep -E "(HTTP|hbase|hdfs|hive|httpfs|hue|impala|mapred|solr|oozie|yarn|zookeeper)[^/]*/[^/]*@" > cluster-princ.txt
             ```

             Open cluster-princ.txt and remove any noncluster service principal entries. Make sure that the default krbtgt and other principals you created, or that were created by Kerberos by default, are not removed by running the following: for i in `cat cluster-princ.txt`; do yes yes | kadmin.local -q "delprinc $i"; done.
         - For an Active Directory KDC, an AD administrator must manually delete the principals for the old hostname from Active Directory.
      2. Start the Cloudera Manager database and Cloudera Manager Server.
      3. Start the Cloudera Manager Agents on the newly renamed hosts. The Agents should show a current heartbeat in Cloudera Manager.

4. Within the Cloudera Manager Admin Console click the Hosts tab.
5. Select the checkbox next to the host with the new name.
6. Select ActionsRegenerate Keytab.

9. If one of the hosts that was renamed has a NameNode configured with high availability and automatic failover enabled, reconfigure the ZooKeeper Failover Controller znodes to reflect the new hostname.
   a) Start ZooKeeper Servers.

      Warning: All other services, and most importantly HDFS, and the ZooKeeper Failover Controller (FC) role within the HDFS, should not be running.

   b) On one of the hosts that has a ZooKeeper Server role, run zookeeper-client.

      1. If the cluster is configured for Kerberos security, configure ZooKeeper authorization as follows:
         - Go to the HDFS service.
         - Click the Instances tab.
         - Click the Failover Controller role.
         - Click the Process tab.
         - In the Configuration Files column of the hdfs/hdfs.sh ["zkfc"] program, expand Show.
         - Inspect core-site.xml in the displayed list of files and determine the value of the ha.zookeeper.auth property, which will be something like: digest:hdfs-fcs:TEbW2bgoODa96rO3ZTn7ND5fSOGx0h. The part after digest:hdfs-fcs: is the password (in the example it is TEbW2bgoODa96rO3ZTn7ND5fSOGx 0h)
         - Run the addauth command with the password:

           ```
           addauth digest hdfs-fcs:TEbW2bgoODa96rO3ZTn7ND5fSOGx0h
           ```

      2. Verify that the HA znode exists: ls /hadoop-ha.
      3. Delete the HDFS znode: rmr /hadoop-ha/nameservice1.
      4. If you are not running JobTracker in a high availability configuration, delete the HA znode: rmr /hadoop-ha.

   c) In the Cloudera Manager Admin Console, go to the HDFS service.
   d) Click the Instances tab.
   e) Select ActionsInitialize High Availability State in ZooKeeper....

10. Update the Hive metastore:
    a) Back up the Hive metastore database.
    b) In the Cloudera Manager Admin Console, go to the Hive service.
    c) Select ActionsUpdate Hive Metastore NameNodes and confirm the command.
11. Update the Database Hostname property for each of the cluster roles for which a database is located on the host being renamed. This is required for both Cloudera Management Service roles (Reports Manager, Activity Monitor, Navigator Audit and Metadata Server) and for cluster services such as Hue, Hive, and so on.
12. Start all cluster services.
13. Start the Cloudera Management Service.
14. Deploy client configurations.

Moving a Host Between Clusters

To move a host between clusters, you must first decommission the host, remove roles from the host, and complete other tasks.

About this task

Important: While you are moving some hosts from one cluster to another (across different Cloudera Manager versions or making any Security configurations), make sure you verify and clear the default paths for the Cloudera Manager Agent install to ensure that old files are not present.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. Decommission the host.
2. Remove all roles from the host (except for the Cloudera Manager management roles).
3. Remove the host from the cluster but leave it available to Cloudera Manager.
4. Add the host to the new cluster.
5. Add roles to the host (optionally using one of the host templates associated with the new cluster).

Configuring Upgrade Domains

Upgrade Domains allow to group cluster hosts for optimal performance during restarts and upgrades.

Upgrade Domains enable faster cluster restarts, faster Cloudera Runtime upgrades, and seamless OS patching & hardware upgrades across large clusters. Upgrade Domains provide an alternative to the default HDFS block placement policy, distributing data across a set of hosts (potentially larger than a single rack) that Cloudera Manager can upgrade/restart at once without compromising service and data availability. When you select Upgrade Domains as the block placement policy, you also assign an Upgrade Domain group to each DataNode host. The NameNode uses these groups to distribute blocks when writing data, and to orchestrate rolling restarts and upgrades. This feature is useful for very large clusters, or for clusters where rolling restarts happen frequently.

For example, if HDFS is configured with the default replication factor of 3, the NameNode places the replica blocks on DataNode hosts in 3 different Upgrade Domains and on at least two different racks.

Note:

- Cloudera recommends that you assign an approximately equal number of DataNode hosts to each Upgrade Domain.
- The number of Upgrade Domains in a cluster should be greater than or equal to the HDFS Replication Factor. When you perform a rolling restart on a cluster, all hosts in an Upgrade Domain group will be restarted simultaneously, followed by the hosts in each remaining Upgrade Domain group.
- You should create a sufficient number of Upgrade Domains so that the cluster can still function adequately when all the hosts in a single Upgrade Domain are taken offline. The appropriate number of Upgrade Domains depends on the workloads and capacity of the cluster and may require tuning for optimal performance.
- To take advantage of the improved rolling restart performance, Upgrade Domain groups should not duplicate rack assignments. The number of hosts in an Upgrade Domain group should be larger than the number of hosts in a rack.

Configuring Upgrade Domains

Steps to configure Upgrade Domains.

About this task

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

Procedure

1. Configure the Upgrade Domains for all hosts:
   a) Click HostsAll Hosts.

b) Select the hosts you want to add to an Upgrade Domain.
c) Click Actions for SelectedAssign Upgrade Domain
d) Enter the name of the Upgrade Domain in the New Upgrade Domain field.
e) Click the Confirm button.

2. Set the HDFS Block Replica Placement Policy:
   a) Open the Cloudera Manager Admin Console.
   b) Go to the HDFS service for the cluster.
   c) Click the Configuration tab.
   d) Search for the HDFS Block Replica Placement Policy configuration parameter.
   e) Select Upgrade Domains.
   f) Click Save Changes.

   The Upgrade Domain assigned to each host displays in the Upgrade Domain column on the All Hosts page. (You may need to add this column to the table: Click the Columns drop-down list above the table and select the Upgrade Domain column.)
3. Restart the HDFS service.

Changing the Upgrade Domain for hosts

Steps to add or change the Upgrade Domain for cluster hosts.

About this task

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

Procedure

1. Click HostsAll Hosts.
2. Select the hosts for the new Upgrade Domain name.
3. Click Actions for SelectedAssign Upgrade Domain
4. Enter the name of the new Upgrade Domain in the New Upgrade Domain field.
5. Click the Confirm button.

Putting all Hosts in an Upgrade Domain group into Maintenance Mode

Steps to put hosts in an Upgrade Domain into Maintenance Mode.

About this task

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

Procedure

1. In Cloudera Manager, select the cluster where you want to decommission hosts.
2. Click HostsAll Hosts.
3. In the Filters section, click Upgrade Domain.
4. Select an Upgrade Domain.
   The All Hosts list now displays only the hosts belonging to the Upgrade Domain.
5. Select all of the hosts.
6. Click Actions for SelectedBegin Maintenance (Suppress Alerts/Decommission.
   The Begin Maintenance (Suppress Alerts/Decommission) dialog box opens. The role instances running on the hosts display at the top. You can also use this dialog box to decommission the host.
7. Select the Take DataNode offline option to put the hosts into Maintenance Mode.

In this mode, alerts from the hosts are suppressed until the host exits Maintenance Mode. The events, however, are still logged. Hosts that are currently in Maintenance Mode display the icon.

8. Click Begin Maintenance.
   The Host Decommission Command dialog box opens and displays the progress of the command.

Specifying Racks for Hosts

To get maximum performance, it is important to configure Cloudera Manager so that it knows the topology of your network. Network locations such as hosts and racks are represented in a tree, which reflects the network "distance" between locations. HDFS will use the network location to place block replicas more intelligently to trade off performance and resilience.

About this task

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

When placing jobs on hosts, CDP prefers within-rack transfers (where there is more bandwidth available) to off-rack transfers; the MapReduce and YARN schedulers use network location to determine where the closest replica is as input to a map task. These computations are performed with the assistance of rack awareness scripts.

Cloudera Manager includes internal rack awareness scripts, but you must specify the racks where the hosts in your cluster are located. If your cluster contains more than 10 hosts, Cloudera recommends that you specify the rack for each host. HDFS, MapReduce, and YARN will automatically use the racks you specify.

Cloudera Manager supports nested rack specifications. For example, you could specify the rack /rack3, or /gro up5/rack3 to indicate the third rack in the fifth group. All hosts in a cluster must have the same number of path components in their rack specifications.

Procedure

1. Click HostsAll Hosts.
2. Select the hosts that you want to assign to a rack.
3. Click Actions for SelectedAssign Rack.
4. Enter a rack name or ID that starts with a slash /, such as /rack123 or /aisle1/rack123.
5. Click Confirm.
6. Optionally, restart any affected services. Rack assignments are not automatically updated for running services.

Performing Maintenance on a Cluster Host

You can perform minor maintenance on cluster hosts by using Cloudera Manager to manage the host decommission and recommission process.

In this process, you can specify whether to suppress alerts from the decommissioned host and, for hosts running the DataNode role, you can specify whether or not to replicate under-replicated data blocks to other DataNodes to maintain the cluster's replication factor. This feature is useful when performing minor maintenance on cluster hosts, such as adding memory or changing network cards or cables where the maintenance window is expected to be short and the extra cluster resources consumed by replicating missing blocks is undesirable.

You can also place hosts into Maintenance Mode, which suppresses unneeded alerts during a maintenance window but does not decommission the hosts.

To perform host maintenance on cluster hosts:

1. Decommission the hosts.
2. Perform the necessary maintenance on the hosts.
3. Recommission the hosts.

Decommissioning Hosts

Cloudera Manager manages the host decommission and recommission process and allows you the option to specify whether to replicate the data to other DataNodes, and whether or not to suppress alerts.

About this task

Decommissioning a host decommissions and stops all roles on the host without requiring you to individually decommission the roles on each service. Decommissioning applies only to the following service roles, other roles running on the host are stopped:

- HDFS DataNode
- MapReduce TaskTracker
- YARN NodeManager
- HBase RegionServer
- Kafka Broker

  Decommissioning only applies to the Kafka Broker role if Cruise Control is installed on the cluster. When a Kafka Broker role is decommissioned, Cruise Control moves the data and partitions from the selected broker role to other broker roles in the cluster. For more information about decommissioning Kafka brokers, see Scaling Kafka brokers.

Hosts with DataNodes and DataNode roles themselves can only be decommissioned if the resulting action leaves enough DataNodes commissioned to maintain the configured HDFS replication factor (by default 3). If you attempt to decommission a DataNode or a host with a DataNode in such situations, the decommission process will not complete and must be aborted.

Before you begin

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

Procedure

To decommission one or more hosts:

1. If the host has a DataNode, and you are planning to replicate data to other hosts (for longer term maintenance operations or to permanently decommission or repurpose the host), perform the steps in Tuning HDFS Prior to Decommissioning DataNodes.
2. In Cloudera Manager, select the cluster where you want to decommission hosts.
3. In the left menu, click HostsAll Hosts.
4. Select the hosts that you want to decommission.
5. Select Actions for SelectedBegin Maintenance (Suppress Alerts/Decommission.

   (If you are logged in as a user with the Limited Operator or Operator role, the menu item is labeled Decommission Host(s) and you will not see the option to suppress alerts.)

   The Begin Maintenance (Suppress Alerts/Decommission) dialog box opens. The role instances running on the hosts display at the top.
6. To decommission the hosts and suppress alerts, select Decommission Host(s). When you select this option for hosts running a DataNode role, choose one of the following (if the host is not running a DataNode role, you will only see the Decommission Host(s) option:):

   - Decommission DataNodes

     This option re-replicates data to other DataNodes in the cluster according to the configured replication factor. Depending on the amount of data and other factors, this can take a significant amount of time and uses a great deal of network bandwidth. This option is appropriate when replacing disks, repurposing hosts for non-HDFS use, or permanently retiring hardware.
   - Take DataNode Offline

This option does not re-replicate HDFS data to other DataNodes until the amount of time you specify has passed, making it less disruptive to active workloads. After this time has passed, the DataNode is automatically recommissioned, but the DataNode role is not started. This option is appropriate for short-term maintenance tasks such not involving disks, such as rebooting, CPU/RAM upgrades, or switching network cables.

Caution: Taking multiple DataNodes offline simultaneously increases the chances that some HDFS data may become unavailable during maintenance. Configuring the proper value for the Maintenance State Minimal Block Replication HDFS configuration property will avoid risking data availability.

Important: Taking DataNode offline supports both traditional 3-replica blocks and Erasure Coding (EC).

7. Click Begin Maintenance.
   The Host Decommission Command dialog box opens and displays the progress of the command.

Results

Note:

- You cannot start roles on a decommissioned host.
- When a DataNode is decommissioned, although HDFS data is replicated to other DataNodes, local files containing the original data blocks are not automatically removed from the storage directories on the host. If you want to permanently remove these files from the host to reclaim disk space, you must do so manually.

What to do next

Perform the necessary maintenance on the hosts.

Recommissioning Hosts

About this task

Only hosts that are decommissioned using Cloudera Manager can be recommissioned.

Before you begin

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

Procedure

1. In Cloudera Manager, select the cluster where you want to recommission hosts.
2. In the left menu, click HostsAll Hosts.
3. Select the hosts that you want to recommission.
4. Select Actions for SelectedEnd Maintenance (Suppress Alerts/Decommission.
   The End Maintenance (Suppress Alerts/Decommission dialog box opens. The role instances running on the hosts display at the top.
5. To recommission the hosts, select Recommission Host(s).
6. Choose one of the following:

   - Bring hosts online and start all roles

     All decommissioned roles will be recommissioned and started. HDFS DataNodes will be started first and brought online before decommissioning to avoid excess replication.
   - Bring hosts online

     All decommissioned roles will be recommissioned but remain stopped. You can restart the roles later.

7. Click End Maintenance.

Results

The Recommission Hosts and Start Roles Command dialog box opens and displays the progress of recommissioning the hosts and restarting the roles

Tuning and Troubleshooting Host Decommissioning

Decommissioning a host decommissions and stops all roles on the host without requiring you to individually decommission the roles on each service. The decommissioning process can take a long time and uses a great deal of cluster resources, including network bandwidth. You can tune the decommissioning process to improve performance and mitigate the performance impact on the cluster.

You can use the Decommission and Recommission features to perform minor maintenance on cluster hosts using Cloudera Manager to manage the process.

Tuning HDFS Prior to Decommissioning DataNodes

When a DataNode is decommissioned, the NameNode ensures that every block from the DataNode will still be available across the cluster as dictated by the replication factor. This procedure involves copying blocks from the DataNode in small batches. If a DataNode has thousands of blocks, decommissioning can take several hours. Before decommissioning hosts with DataNodes, you should first tune HDFS:

About this task

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

Procedure

1. Run the following command to identify any problems in the HDFS file system:

```
hdfs fsck / -list-corruptfileblocks -openforwrite -files -blocks -locati
ons 2>&1 > /tmp/hdfs-fsck.txt
```

2. Fix any issues reported by the fsck command. If the command output lists corrupted files, use the fsck command to move them to the lost+found directory or delete them:

```
hdfs fsck FILE_NAME -move
```

or

```
hdfs fsck FILE_NAME -delete
```

3. Raise the heap size of the DataNodes. DataNodes should be configured with at least 4 GB heap size to allow for the increase in iterations and max streams.
   a) Go to the HDFS service page.
   b) Click the Configuration tab.
   c) Select ScopeDataNode.
   d) Select CategoryResource Management.
   e) Set the Java Heap Size of DataNode in Bytes property as recommended.

      To apply this configuration property to other role groups as needed, edit the value for the appropriate role group.
4. Increase the Replication Work Multiplier Per Iteration to a larger number. The default value is 10. Change the value to a larger value between 20 to 30.
   a) Select Scope NameNode .
   b) Expand the Category Advanced category.

c) Configure the Replication Work Multiplier Per Iteration property to a larger number.

To apply this configuration property to other role groups as needed, edit the value for the appropriate role group.

5. Increase the replication maximum threads and maximum replication thread hard limits.
   a) Select ScopeNameNode.
   b) Expand the CategoryAdvanced category.
   c) Configure the Maximum number of replication threads on a DataNode and Hard limit on the number of replication threads on a DataNode properties to 50 and 100 respectively. You can decrease the number of threads (or use the default values) to minimize the impact of decommissioning on the cluster, but the trade off is that decommissioning will take longer.

   To apply this configuration property to other role groups as needed, edit the value for the appropriate role group.
6. Restart the HDFS service.

Related Information

Performance Considerations

Modifying Configuration Properties Using Cloudera Manager

Tuning HBase Prior to Decommissioning DataNodes

To increase the speed of a rolling restart of the HBase service, set the Region Mover Threads property to a higher value.

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

This increases the number of regions that can be moved in parallel, but places additional strain on the HMaster. In most cases, Region Mover Threads should be set to 5 or lower.

Performance Considerations

Decommissioning a DataNode does not happen instantly because the process requires replication of a potentially large number of blocks. During decommissioning, the performance of your cluster may be impacted.

This section describes the decommissioning process and suggests solutions for several common performance issues.

Decommissioning occurs in two steps:

1. The Commission State of the DataNode is marked as Decommissioning and the data is replicated from this node to other available nodes. Until all blocks are replicated, the node remains in a Decommissioning state. You can view this state from the NameNode Web UI. (Go to the HDFS service and select Web UINameNode Web UI.)
2. When all data blocks are replicated to other nodes, the node is marked as Decommissioned.

Decommissioning can impact performance in the following ways:

- There must be enough disk space on the other active DataNodes for the data to be replicated. After decommissioning, the remaining active DataNodes have more blocks and therefore decommissioning these DataNodes in the future may take more time.
- There will be increased network traffic and disk I/O while the data blocks are replicated.
- Data balance and data locality can be affected, which can lead to a decrease in performance of any running or submitted jobs.
- Decommissioning a large numbers of DataNodes at the same time can decrease performance.
- If you are decommissioning a minority of the DataNodes, the speed of data reads from these nodes limits the performance of decommissioning because decommissioning maxes out network bandwidth when reading data blocks from the DataNode and spreads the bandwidth used to replicate the blocks among other DataNodes in the cluster. To avoid performance impacts in the cluster, Cloudera recommends that you only decommission a minority of the DataNodes at the same time.
- You can decrease the number of replication threads to decrease the performance impact of the replications, but this will cause the decommissioning process to take longer to complete.

Cloudera recommends that you add DataNodes and decommission DataNodes in parallel, in smaller groups. For example, if the replication factor is 3, then you should add two DataNodes and decommission two DataNodes at the same time.

Related Information

Tuning HDFS Prior to Decommissioning DataNodes

Troubleshooting Performance of Decommissioning

Several conditions can impact performance when you decommission DataNodes.

Open Files

Write operations on the DataNode do not involve the NameNode. If there are blocks associated with open files located on a DataNode, they are not relocated until the file is closed. This commonly occurs with:

- Clusters using HBase
- Open Flume files
- Long running tasks

To find open files, run the following command:

```
hdfs dfsadmin -listOpenFiles -blockingDecommission
```

The command returns output similar to the following example:

```
Client Host         Client Name         Open File Path
172.26.12.77        DFSClient_NONMAPREDUCE_-698274460_1 /hbase/ol
dWALs/dn3.cloudera.com%2C22101%2C1540973344249.dn3.cloudera.com%
2C22101%2C1540973344249.regiongroup-0.154099857098
```

After you find the open files, perform the appropriate action to restart process to close the file. For example, major compaction closes all files in a region for HBase.

Alternatively, you may evict writers to those decommissioning DataNodes with the following command:

```
hdfs dfsadmin -evictWriters <datanode_host:ipc_port>
```

For example:

```
hdfs dfsadmin -evictWriters datanode1:20001
```

A block cannot be relocated because there are not enough DataNodes to satisfy the block placement policy.

For example, for a 10 node cluster, if the mapred.submit.replication is set to the default of 10 while attempting to decommission one DataNode, there will be difficulties relocating blocks that are associated with map/reduce jobs. This condition will lead to errors in the NameNode logs similar to the following:

```
org.apache.hadoop.hdfs.server.blockmanagement.BlockPlacementPoli
cyDefault: Not able to place enough replicas, still in need of 3
 to reach 3
```

Use the following steps to find the number of files where the block replication policy is equal to or above your current cluster size:

1. Provide a listing of open files, their blocks, the locations of those blocks by running the following command:

```
hadoop fsck / -files -blocks -locations -openforwrite 2>&1 >
 openfiles.out
```

2. Run the following command to return a list of how many files have a given replication factor:

```
grep repl= openfiles.out | awk '{print $NF}' | sort | uniq -c
```

   For example, when the replication factor is 10 , and decommissioning one:

```
egrep -B4 "repl=10" openfiles.out | grep -v '<dir>' | awk '/^
\//{print $1}'
```

3. Examine the paths, and decide whether to reduce the replication factor of the files, or remove them from the cluster.

Maintenance Mode

Maintenance mode allows you to suppress alerts for a host, service, role, or an entire cluster. This can be useful when you need to take actions in your cluster (make configuration changes and restart various elements) and do not want to see the alerts that will be generated due to those actions.

Putting an entity into maintenance mode does not prevent events from being logged; it only suppresses the alerts that those events would otherwise generate. You can see a history of all the events that were recorded for entities during the period that those entities were in maintenance mode.

Explicit and Effective Maintenance Mode

When you enter maintenance mode on an entity (cluster, service, or host) that has subordinate entities (for example, the roles for a service) the subordinate entities are also put into maintenance mode. These are considered to be in *effective maintenance mode*, as they have inherited the setting from the higher-level entity.

For example:

- If you set the HBase service into maintenance mode, then its roles (HBase Master and all RegionServers) are put into effective maintenance mode.
- If you set a host into maintenance mode, then any roles running on that host are put into effective maintenance mode.

Entities that have been explicitly put into maintenance mode show the icon . Entities that have entered effective maintenance mode as a result of inheritance from a higher-level entity show the icon .

When an entity (role, host or service) is in effective maintenance mode, it can only be removed from maintenance mode when the higher-level entity exits maintenance mode. For example, if you put a service into maintenance mode, the roles associated with that service are entered into effective maintenance mode, and remain in effective maintenance mode until the service exits maintenance mode. You cannot remove them from maintenance mode individually.

Alternatively, an entity that is in effective maintenance mode can be put into explicit maintenance mode. In this case, the entity remains in maintenance mode even when the higher-level entity exits maintenance mode. For example, suppose you put a host into maintenance mode, (which puts all the roles on that host into effective maintenance mode). You then select one of the roles on that host and put it explicitly into maintenance mode. When you have the host exit maintenance mode, that one role remains in maintenance mode. You need to select it individually and specifically have it exit maintenance mode.

Entering Maintenance Mode

You can enable maintenance mode for a cluster, service, role, or host.

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

Putting a Cluster into Maintenance Mode

1. In the left menu, click Clusters<cluster name>.
2. Click the Actions menu ( ) to the right of the cluster name and select Enter Maintenance Mode.
3. Confirm that you want to do this.

The cluster is put into explicit maintenance mode, as indicated by the icon. All services and roles in the cluster are entered into effective maintenance mode, as indicated by the icon.

Putting a Service into Maintenance Mode

1. In the left menu, click Clusters and select the service.
2. Click ActionsEnter Maintenance Mode.
3. Confirm that you want to do this.

The service is put into explicit maintenance mode, as indicated by the icon. All roles for the service are entered into effective maintenance mode, as indicated by the icon.

Putting Roles into Maintenance Mode

1. In the left menu, click Clusters and select the service.
2. Click the Instances tab.
3. Select the role(s) you want to put into maintenance mode.
4. From the Actions for Selected menu, select Enter Maintenance Mode.
5. Confirm that you want to do this.

The roles will be put in explicit maintenance mode. If the roles were already in effective maintenance mode (because its service or host was put into maintenance mode) the roles will now be in explicit maintenance mode. This means that they will not exit maintenance mode automatically if their host or service exits maintenance mode; they must be explicitly removed from maintenance mode.

Putting Hosts into Maintenance Mode

1. In Cloudera Manager, select the cluster where you want to decommission hosts.
2. Click HostsAll Hosts.
3. Select the hosts that you want to put into Maintenance Mode.
4. Select Actions for SelectedBegin Maintenance (Suppress Alerts/Decommission.

   The Begin Maintenance (Suppress Alerts/Decommission) dialog box opens. The role instances running on the hosts display at the top. You can also use this dialog box to decommission the host.
5. Deselect the Decommission Host(s) option to put the host into Maintenance Mode. In this mode, alerts from the hosts are suppressed until the host exits Maintenance Mode. The events, however, are still logged. Hosts that are currently in Maintenance Mode display the icon.
6. Click Begin Maintenance.

   The Host Decommission Command dialog box opens and displays the progress of the command.

Exiting Maintenance Mode

When you exit maintenance mode, the maintenance mode icons are removed and alert notification resumes.

Exiting a Cluster from Maintenance Mode

1. Click to the right of the cluster name and select Exit Maintenance Mode.
2. Confirm that you want to do this.

Exiting a Service from Maintenance Mode

1. Click to the right of the service name and select Exit Maintenance Mode.
2. Confirm that you want to do this.

Exiting Roles from Maintenance Mode

1. Go to the services page that includes the role.
2. Go to the Instances tab.
3. Select the role(s) you want to exit from maintenance mode.
4. From the Actions for Selected menu, select Exit Maintenance Mode.
5. Confirm that you want to do this.

Taking Hosts out of Maintenance Mode

1. In Cloudera Manager, to go the cluster with the hosts you want to take out of Maintenance Mode.
2. Click HostsAll Hosts.
3. Select the hosts that are ready to exit Maintenance Mode.
4. Select Actions for SelectedEnd Maintenance (Suppress Alerts/Decommission.

   The End Maintenance (Suppress Alerts/Decommission) dialog box opens. The role instances running on the hosts display at the top.
5. Deselect the Recommission Host(s) option to take the host out of Maintenance Mode and re-enable alerts from the hosts. Hosts that are currently in Maintenance Mode display the  icon on the All Hosts page.
6. Click End Maintenance.

Viewing the Maintenance Mode Status of a Cluster

For any cluster, you can view the components (service, roles, or hosts) that are in maintenance mode.

About this task

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

Procedure

1. From the Cloudera Manager Home page, select the cluster that you want to view the maintenance mode status for.
2. Click Actions View Maintenance Mode Status... .
   This pops up a dialog box that shows the components in your cluster that are in maintenance mode, and indicates which are in effective maintenance mode as well as those that have been explicitly placed into maintenance mode.

   From this dialog box you can select any of the components shown there and remove them from maintenance mode.

   If individual services are in maintenance mode, you will see the maintenance mode icon next to the Actions button for that service.

   Note: The Actions button is not enabled if you are viewing status for a point of time in the past.

Managing Roles

When Cloudera Manager configures a service, it configures hosts in your cluster with one or more functions (called roles in Cloudera Manager) that are required for that service. Understand how you can manage roles.

Roles determine which Hadoop daemons run on a given host. For example, when Cloudera Manager configures an HDFS service instance it configures one host to run the NameNode role, another host to run as the Secondary NameNode role, another host to run the Balancer role, and some or all of the remaining hosts to run DataNode roles.

Configuration settings are organized in role groups. A *role group* includes a set of configuration properties for a specific group, as well as a list of role instances associated with that role group. Cloudera Manager automatically creates default role groups.

For role types that allow multiple instances on multiple hosts, such as DataNodes, TaskTrackers, RegionServers (and many others), you can create multiple role groups to allow one set of role instances to use different configuration settings than another set of instances of the same role type. In fact, upon initial cluster setup, if you are installing on identical hosts with limited memory, Cloudera Manager will (typically) automatically create two role groups for each worker role — one group for the role instances on hosts with only other worker roles, and a separate group for the instance running on the host that is also hosting master roles.

The HDFS service is an example of this: Cloudera Manager typically creates one role group (DataNode Default Group) for the DataNode role instances running on the worker hosts, and another group (HDFS-1-DATANODE-1) for the DataNode instance running on the host that is also running the master roles such as the NameNode, JobTracker, HBase Master and so on. Typically the configurations for those two classes of hosts will differ in terms of settings such as memory for JVMs.

Cloudera Manager configuration screens offer two layout options: classic and new. The new layout is the default; however, on each configuration page you can easily switch between layouts using the Switch to *XXX* layout link at the top right of the page.

Gateway Roles

A *gateway* is a special type of role whose sole purpose is to designate a host that should receive a client configuration for a specific service, when the host does not have any roles running on it. Gateway roles enable Cloudera Manager to install and manage client configurations on that host. There is no process associated with a gateway role, and its status will always be Stopped. You can configure gateway roles for HBase, HDFS, Hive, Kafka, MapReduce, Solr, Spark, Sqoop 1 Client, and YARN.



Warning:

Gateway roles are lightweight, but excessive distribution of them at large scale can negatively affect Cloudera Manager Server performance.

Related Information

Cluster Configuration Overview

Role Instances

Adding a Role Instance

About this task

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

After creating services, you can add role instances to the services. For example, after initial installation in which you created the HDFS service, you can add a DataNode role instance to a host where one was not previously running.

Upon upgrading a cluster to a new version of Cloudera Runtime you might want to create a role instance for a role added in the new version.

Procedure

1. Go to the service for which you want to add a role instance. For example, to add a DataNode role instance, go to the HDFS service.
2. Click the Instances tab.
3. Click the Add Role Instances button.
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. The Assign Templates page displays. Click the drop-down list for the role you want to add and select a Host Template.
6. Click Continue.
7. In the Review Changes page, review the configuration changes to be applied.

   Confirm the settings entered for file system paths. The file paths required vary based on the services to be installed. For example, you might confirm the NameNode Data Directory and the DataNode Data Directory for HDFS.
8. Click Continue.

Results

The wizard finishes by performing any actions necessary to prepare the cluster for the new role instances. For example, new DataNodes are added to the NameNode dfs_hosts_allow.txt file. The new role instance is configured with the default role group for its role type, even if there are multiple role groups for the role type. If you want to use a different role group, follow the instructions in the topic *Managing Role Groups* for moving role instances to a different role group

Related Information

Managing Role Groups

Starting, Stopping, and Restarting Role Instances

About this task

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

If the host for the role instance is currently decommissioned, you will not be able to start the role until the host has been recommissioned.

Important: Use Cloudera Manager to stop the Node Manager service. If it is stopped manually, it can cause jobs to fail.

Procedure

1. Go to the service that contains the role instances to start, stop, or restart.
2. Click the Instances tab.
3. Check the checkboxes next to the role instances to start, stop, or restart (such as a DataNode instance).
4. Select Actions for SelectedStart, Stop, or Restart, and then click Start, Stop, or Restart again to start the process. When you see a Finished status, the process has finished.

Related Information

Rolling Restart

Decommissioning Role Instances

You can remove a role instance such as a DataNode from a cluster while the cluster is running by decommissioning the role instance.

About this task

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

When you decommission a role instance, Cloudera Manager performs a procedure so that you can safely retire a host without losing data. Role decommissioning applies to the following service roles:

- HDFS DataNode
- MapReduce TaskTracker
- YARN NodeManager
- HBase RegionServer
- Kafka Broker

  Decommissioning only applies to the Kafka Broker role if Cruise Control is installed on the cluster. When a Kafka Broker role is decommissioned, Cruise Control moves the data and partitions from the selected broker role to other broker roles in the cluster. For more information about decommissioning Kafka brokers, see Scaling Kafka brokers.

Hosts with DataNodes and DataNode roles themselves can only be decommissioned if the resulting action leaves enough DataNodes commissioned to maintain the configured HDFS replication factor (by default 3). If you attempt to decommission a DataNode or a host with a DataNode in such situations, the decommission process will not complete and must be aborted.

A role will be decommissioned if its host is decommissioned.

To remove a DataNode from the cluster, you decommission the DataNode role as described here and then perform a few additional steps to remove the role. See the topic Delete a DataNode.

Procedure

To decommission role instances:

1. If you are decommissioning DataNodes, perform the steps in the topic *Tuning HDFS Prior to Decommissioning DataNodes*.
2. Click the service instance that contains the role instance you want to decommission.
3. Click the Instances tab.
4. Check the checkboxes next to the role instances to decommission.
5. Select Actions for SelectedDecommission, and then click Decommission again to start the process.

Results

A Decommission Command pop-up displays that shows each step or decommission command as it is run. In the Details area, click › to see the subcommands that are run. Depending on the role, the steps may include adding the host to an "exclusions list" and refreshing the NameNode, JobTracker, or NodeManager; stopping the Balancer (if it is running); and moving data blocks or regions. Roles that do not have specific decommission actions are stopped.

You can abort the decommission process by clicking the Abort button, but you must recommission and restart the role.

The Commission State facet in the Filters list displays Decommissioning while decommissioning is in progress, and Decommissioned when the decommissioning process has finished. When the process is complete, a ✓ is added in front of Decommission Command.

Related Information

Tuning HDFS Prior to Decommissioning DataNodes

Recommissioning Role Instances

About this task

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

Procedure

1. Click the service that contains the role instance you want to recommission.
2. Click the Instances tab.
3. Check the checkboxes next to the decommissioned role instances to recommission.
4. Select Actions for SelectedRecommission, and then click Recommission to start the process. A Recommission Command pop-up displays that shows each step or recommission command as it is run. When the process is complete, a ✓ is added in front of Recommission Command.
5. Restart the role instance.

Deleting Role Instances

About this task

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

Deleting Role Instances

Procedure

1. Click the service instance that contains the role instance you want to delete. For example, if you want to delete a DataNode role instance, click an HDFS service instance.
2. Click the Instances tab.
3. Check the checkboxes next to the role instances you want to delete.
4. If the role instance is running, select Actions for SelectedStop and click Stop to confirm the action.
5. Select Actions for SelectedDelete. Click Delete to confirm the deletion.

Results

Note: Deleting a role instance does not clean up the associated client configurations that have been deployed in the cluster.

Deleting Role Instances from a Host Template

About this task

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

Before you begin

Note that role instances cannot be directly deleted from hosts. All the roles within a role group must be deleted together. Removing any role group from the template will delete the associated roles. Follow the steps below to delete role instances.

Procedure

1. Go to HostsHost Templates.
   The Host Templates page displays.
2. Click the action menu on the right in the row containing the role you want to delete and select Edit.
3. Expand the row for the role's service.
4. De-select the role group(s) you want to delete.
5. Click Save And Re-apply.

Results



Note: Deleting a role instance does not clean up the associated client configurations that have been deployed in the cluster. You may need to restart the service and/or re-deploy client configurations after modifying the template.

Configuring Roles to Use a Custom Garbage Collection Parameter

You can use Java configuration options to configure roles to use a custom garbage collection parameter.

Every Java-based role in Cloudera Manager has a configuration setting called Java Configuration Options for *ROLE* where you can enter command line options. Commonly, garbage collection flags or extra debugging flags would be passed here. To find the appropriate configuration setting, select the service you want to modify in the Cloudera Manager Admin Console, then use the Search box to search for Java Configuration Options.

You can add configuration options for all instances of a given role by making this configuration change at the service level. For example, to modify the setting for all DataNodes, select the HDFS service, then modify the Java Configuration Options for DataNode setting.

To modify a configuration option for a given instance of a role, select the service, then select the particular role instance (for example, a specific DataNode). The configuration settings you modify will apply to the selected role instance only.

Related Information

Modifying Configuration Properties Using Cloudera Manager

Role Groups

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

A *role group* is a set of configuration properties for a role type, as well as a list of role instances associated with that group. Cloudera Manager automatically creates a default role group named *ROLE TYPE* Default Group for each role type.Each role instance can be associated with only a single role group.

Role groups provide two types of properties: those that affect the configuration of the service itself and those that affect monitoring of the service, if applicable (the Monitoring subcategory). Not all services have monitoring properties.

When you run the installation or upgrade wizard, Cloudera Manager configures the default role groups it adds, and adds any other required role groups for a given role type. For example, a DataNode role on the same host as the NameNode might require a different configuration than DataNode roles running on other hosts. Cloudera Manager creates a separate role group for the DataNode role running on the NameNode host and uses the default configuration for DataNode roles running on other hosts.

You can modify the settings of the default role group, or you can create new role groups and associate role instances to whichever role group is most appropriate. This simplifies the management of role configurations when one group of role instances may require different settings than another group of instances of the same role type—for example, due to differences in the hardware the roles run on. You modify the configuration for any of the service's role groups through the Configuration tab for the service. You can also override the settings inherited from a role group for a role instance.

If there are multiple role groups for a role type, you can move role instances from one group to another. When you move a role instance to a different group, it inherits the configuration settings for its new group.

Related Information

Configuring Monitoring Settings
Overriding Configuration Properties

Creating a Role Group

About this task

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

Procedure

1. Go to a service status page.
2. Click the Instances or Configuration tab.
3. Click Role Groups.
4. Click Create new group....
5. Provide a name for the group.
6. Select the role type for the group. You can select role types that allow multiple instances and that exist for the service you have selected.
7. In the Copy From field, select the source of the basic configuration information for the role group:
   - An existing role group of the appropriate type.
   - None.... The role group is set up with generic default values that are not the same as the values Cloudera Manager sets in the default role group, as Cloudera Manager specifically sets the appropriate configuration properties for the services and roles it installs. After you create the group you must edit the configuration to set missing properties (for example the TaskTracker Local Data Directory List property, which is not populated if you select None) and clear other validation warnings and errors.

Related Information

Modifying Configuration Properties Using Cloudera Manager

Managing Role Groups

About this task

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

Procedure

1. Go to a service status page.
2. Click the Instances or Configuration tab.
3. Click Role Groups.
4. Click the group you want to manage. Role instances assigned to the role group are listed.
5. Perform the appropriate procedure for the action:
   - Rename
     a. Click the role group name, and click Rename.
     b. Specify the new name and click Rename.
   - Delete

     You cannot delete any of the default groups. The group must first be empty; if you want to delete a group you've created, you must move any role instances to a different role group.

     a. Click the role group name.
     b. Click Delete, and confirm by clicking Delete. Deleting a role group removes it from host templates.
   - Move
     a. Select the role instance(s) to move.
     b. Select Actions for SelectedMove To Different Role Group....
     c. In the pop-up that appears, select the target role group and click Move.
   - Move
     a. Go to HostsHost Templates
     b. Click the action menu on the right in the row containing the role you want to move and select Edit.
     c. Modify the role group you want to move by removing the old role group and adding the new role group to the host template.
     d. Click Save.

        All the associated roles are moved to the new role group.

Related Information

Managing Hosts

Default User Roles

By default, Cloudera Manager ships with user roles that have privileges for all clusters managed by Cloudera Manager.

The following table describes the actions each user role can perform:

Permitted Operations	Auditor	Cluster Administrator	Cluster Creator	Configurator	Dashboard User	Full Administrator	Key Administrator	Limited Cluster Administrator	Limited Operator	Navigator Administrator	Operator	Read-Only	Replication Administrator	User Administrator
Access all functionality that Cloudera Manager offers		Y				Y								
Add and Remove Entity Tags		Y				Y		Y						
Administer Cloudera Navigator		Y				Y				Y				
Apply policies to redact sensitive data		Y				Y								
Configure HDFS Encryption, administer Key Trustee Server, and manage encryption keys						Y	Y							
Create clusters		Y	Y			Y								
Create replication policies and snapshot policies						Y							Y	
Create, modify, and delete your own dashboards					Y	Y								
Create, update, or delete external account configuration						Y								Y
Decommission hosts		Y		Y		Y		Y	Y		Y			
Edit the configuration of services and roles		Y		Y		Y		Y						
Enter and exit Maintenance Mode		Y		Y		Y		Y						
Import Cluster Template		Y				Y		Y						
Inspect Hosts		Y				Y		Y						
Manage Full Administrator accounts						Y								
Manage user accounts and configuration of external authentication						Y								Y
Recommission hosts, and decommission and recommission roles		Y		Y		Y		Y			Y			
See available hosts		Y	Y			Y		Y						
Send Diagnostic Bundles		Y				x		Y						
Start, stop, and restart KMS		Y		Y		Y	Y	Y			Y			
Start, stop, and restart most clusters, services, and roles		Y		Y		Y		Y			Y			
Upgrade Clusters		Y				Y								
View and perform parcels operations		Y	Y			Y		Y						
View audit events	Y					Y				Y				
View data in Cloudera Manager	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Disk Usage By Directory	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Directory Usage		Y				Y								
File Browser		Y				Y								
Kill applications in YARN		Y				Y								
Cancel memory-intensive Impala queries		Y				Y								

Related Information

Cloudera Manager User Roles

Backing up Cloudera Manager databases

Cloudera recommends that you schedule regular backups of the databases that Cloudera Manager uses to store configuration, monitoring, and reporting data and for managed services that require a database:

Backing Up PostgreSQL Databases

To back up a PostgreSQL database, use the same procedure whether the database is embedded or external:

1. Log in to the host where the Cloudera Manager Server is installed.
2. Get the name, user, and password properties for the Cloudera Manager database from /etc/cloudera-scm-server/db.properties:

```
com.cloudera.cmf.db.name=scm
com.cloudera.cmf.db.user=scm
com.cloudera.cmf.db.password=NnYfWIjlbk
```

3. Run the following command as root using the parameters from the preceding step:

```
# pg_dump -h HOSTNAME -p 7432 -U scm > /tmp/scm_server_db_backup.$(date +
%Y%m%d)
```

4. Enter the password from the com.cloudera.cmf.db.password property in step 2.

5. To back up a database created for one of the roles on the local host as the *ROLEUSER* user:

```
# pg_dump -h HOSTNAME -p 7432 -U ROLEUSER > /tmp/ROLEDB
```

6. Enter the password specified when the database was created.

Backing Up MariaDB Databases

To back up the MariaDB database, run the mysqldump command on the MariaDB host, as follows:

```
mysqldump -hHOSTNAME -uUSERNAME -pPASSWORD DATABASE > /tmp/DATABASE-
BACKUP.sql
```

For example, to back up the Activity Monitor database amon created in Creating Databases for Cloudera Software, on the local host as the root user, with the password amon_password:

```
mysqldump -pamon_password amon > /tmp/amon-backup.sql
```

To back up the sample Activity Monitor database amon on remote host myhost.example.com as the root user, with the password amon_password:

```
mysqldump -hmyhost.example.com -uroot -pamon_password amon > /tmp/amon-backu
p.sql
```

Backing Up MySQL Databases

To back up the MySQL database, run the mysqldump command on the MySQL host, as follows:m

```
mysqldump -hHOSTNAME -uUSERNAME -pPASSWORD DATABASE > /tmp/DATABASE-
BACKUP.sql
```

For example, to back up the Activity Monitor database amon created in Creating Databases for Cloudera Software, on the local host as the root user, with the password amon_password:

```
mysqldump -pamon_password amon > /tmp/amon-backup.sql
```

To back up the sample Activity Monitor database amon on remote host myhost.example.com as the root user, with the password amon_password:

```
mysqldump -hmyhost.example.com -uroot -pamon_password amon > /tmp/amon-backu
p.sql
```

You can back up all database using the following command:

```
mysqldump --all-databases -pPASSWORD > /tmp/all1/all.sql
```

Backing Up Oracle Databases

For Oracle, work with your database administrator to ensure databases are properly backed up.

Core Settings Service

The Core Configuration service has been enhanced and renamed to Core Settings service. The Core Settings service allows you to create clusters without the HDFS service and is used to store cluster-wide settings, including some settings that were previously stored in HDFS or in the legacy Core Configuration service.

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

Overview

The Core Settings service allows you to create more types of clusters without having to include the HDFS service. Previously, the HDFS service was required in many cases even when data was not being stored in HDFS because some services like Sentry and Spark required cluster-wide configuration files that Cloudera Manager deploys using the HDFS service. The Core Settings service provides this configuration in a standalone fashion and thus eliminates the need for an HDFS service.

Beginning with Cloudera Manager 7.7.1 the Core Settings service is implemented in a new way that changes how it is added to the cluster and what occurs during an upgrade of Cloudera Manager. There are also some potential changes to how the Cloudera Manager API is used to access configuration data stored in the Core Settings service.

New installations

For new installations, the Core Settings service is added to newly created clusters by default.

Upgrades

After you upgrade to Cloudera Manager 7.7.1 (or higher), Cloudera Manager executes several steps when the Cloudera Manager server first starts after the upgrade:

- The Core Settings service is automatically added to clusters managed by Cloudera Manager that did not have the Core Configuration service.
- Some Configurations will be removed from HDFS and stored in the Core Settings service. See Configuration parameters migrated to Core Settings Service on page 70 for a list of these parameters.
- If the cluster already has the legacy Core Configuration service, Cloudera Manager will update this service (but keeping the original service name), and also add a StubDFS service. Further, the Storage Operations role of the Core Configuration Service is moved to the StubDFS service. This service provides backwards compatibility for services that have a mandatory DFS service dependency, allowing access to core configuration files..



Note: After upgrading Cloudera Manager, some services may report stale configurations in the Cloudera Manager Admin Console. Although configurations have not changed, the way they are stored has changed, and therefore Cloudera Manager infers that there have been changes to the files. To clear the stale configurations, restart the service, or the cluster and re-deploy the Client configurations. Where available, you can use the rolling restart option to avoid down time.



Note: The Core Settings service cannot be deleted.

The Cloudera Manager API and the Core Settings service

Calls to the Cloudera Manager API to access configuration settings that previously were owned by HDFS will continue to work as expected without code changes using the "API Compatibility Mode". This mode transparently redirects such API calls to perform the configuration changes on the Core Settings Service and is enabled by default.



Important: Accessing core configuration via HDFS is deprecated and the compatibility mode will be removed from Cloudera Manager in a later release.

If you want to revise your code to access the settings without using the Compatibility mode, change your API consumers to target the Core Settings service instead of HDFS. Cloudera highly recommends that you test and migrate your code by disabling the API Compatibility mode to avoid any future interruptions.

You can test this code by disabling the API Compatibility mode:

1. Open the Cloudera Manager Admin Console.
2. Go to the HDFS service page.
3. Click the Configuration tab.
4. Search for the following configuration parameter: Legacy Cloudera Manager API Clients Compatibility.

Configuration parameters migrated to Core Settings Service

During a Cloudera Manager upgrade, Cloudera Manager migrates the parameters listed on this page to the Core Settings Service.



Note: Once the Cloudera Manager upgrade process is completed, most of the configurations in Isilon/ powerscale are not initialized in the Custom Software Development solutions (CSD) and they seen coming from Distributed File System (DFS) provider. Later, the configurations are removed from DFS provider and added in coresetting as part of hdfsless project. It is seen that some of the configurations are removed from Isilon as well, once the upgrade process is completed.

Currently, there are no upgradehandlers to copy configurations from third party CSD (you can only copy configurations from HDFS to core settings during the upgrade process).

You can employ this workaround to manually modify the appropriate configuration values after the Cloudera Manager upgrade process is completed.

- extra_auth_to_local_rules
- hadoop_authorized_admin_groups
- hadoop_authorized_admin_users
- hadoop_authorized_groups
- hadoop_authorized_users
- hadoop_group_mapping_ldap_base
- hadoop_group_mapping_ldap_bind_passwd
- hadoop_group_mapping_ldap_bind_user
- hadoop_group_mapping_ldap_group_filter
- hadoop_group_mapping_ldap_group_name_attr
- hadoop_group_mapping_ldap_keystore
- hadoop_group_mapping_ldap_keystore_passwd
- hadoop_group_mapping_ldap_member_attr
- hadoop_group_mapping_ldap_url
- hadoop_group_mapping_ldap_use_ssl
- hadoop_group_mapping_ldap_user_filter
- hadoop_http_auth_cookie_domain
- hadoop_rpc_protection
- hadoop_security_authentication
- hadoop_security_authorization
- hadoop_security_group_mapping
- hdfs_encryption_key_length
- hdfs_hadoop_ssl_enabled
- hdfs_proxy_user_groups_list
- hdfs_proxy_user_hosts_list
- hfds_ssl_client_safety_valve
- hive_proxy_user_groups_list

- hive_proxy_user_hosts_list
- http_auth_signature_secret
- HTTP_proxy_user_groups_list
- HTTP_proxy_user_hosts_list
- httpfs_proxy_user_groups_list
- httpfs_proxy_user_hosts_list
- hue_proxy_user_groups_list
- hue_proxy_user_hosts_list
- impala_proxy_user_groups_list
- impala_proxy_user_hosts_list
- knox_proxy_user_groups_list
- knox_proxy_user_hosts_list
- kudu_proxy_user_groups_list
- kudu_proxy_user_hosts_list
- livy_proxy_user_groups_list
- livy_proxy_user_hosts_list
- mapred_proxy_user_groups_list
- mapred_proxy_user_hosts_list
- object_store_service
- oozie_proxy_user_groups_list
- oozie_proxy_user_hosts_list
- phoenix_proxy_user_groups_list
- phoenix_proxy_user_hosts_list
- redaction_policy
- redaction_policy_enabled
- security_logger_enabled
- set_auth_to_local_to_lowercase
- smon_proxy_user_groups_list
- smon_proxy_user_hosts_list
- ssl_client_truststore_location
- ssl_client_truststore_password
- strict_transport_security
- telepub_proxy_user_groups_list
- telepub_proxy_user_hosts_list
- trusted_realms
- yarn_proxy_user_groups_list
- yarn_proxy_user_hosts_list

Performance Management

This section describes mechanisms and best practices for improving performance.

Optimizing Performance in Cloudera Runtime

This section provides solutions to some performance problems, and describes configuration best practices.

Important: Work with your network administrators and hardware vendors to ensure that you have the proper NIC firmware, drivers, and configurations in place and that your network performs properly. Cloudera recognizes that network setup and upgrade are challenging problems, and will do its best to share useful experiences.

Review our CDP Reference Architecture for additional hardware and operating system guidance.

Disabling Transparent Hugepages (THP)

Most Linux platforms supported by Cloudera Runtime include a feature called *transparent hugepages*, which interacts poorly with Hadoop workloads and can seriously degrade performance.

About this task

Symptoms:

Symptom 1: High system CPU usage and performance degradation

System monitoring tools, such as top, show a large percentage of CPU usage under "system CPU." If system CPU usage consumes 30% or more of the total CPU capacity, your system might be experiencing this issue.

To check if the system enables transparent hugepages, run the following commands and verify the output:

```
$ cat DEFRAG_FILE_PATHNAME
$ cat ENABLED_FILE_PATHNAME
```

- [always] never means that transparent hugepages is enabled.
- always [never] means that transparent hugepages is disabled.

Symptom 2: Server native memory leak and 100% CPU spin (OPSAPS-77888)

When you enable Transparent Huge Pages (THP) and set them to always on the cluster, a known issue with the Oracle Database driver (Oracle Instant Client) triggers severe memory instability by causing it to consume a large amount of memory.

The AnonHugePages memory allocation grows continuously, causing a server native memory leak. Consequently, the memory consumption of processes using this database driver, such as Cloudera Manager, keeps growing and escalates until it induces a 100% CPU spin, which eventually prevents all other system processes from functioning.

To disable Transparent Hugepages, perform the following steps on all cluster hosts:

Procedure

1. To disable transparent hugepages on reboot, add the following commands to the /etc/rc.d/rc.local file on all cluster hosts:
   - RHEL/CentOS 7.x, 8.x, and 9.x:

     ```
     echo never > /sys/kernel/mm/transparent_hugepage/enabled
     echo never > /sys/kernel/mm/transparent_hugepage/defrag
     ```

   - RHEL/CentOS 6.x

     ```
     echo never > /sys/kernel/mm/redhat_transparent_hugepage/defrag
     echo never > /sys/kernel/mm/redhat_transparent_hugepage/enabled
     ```

   - Ubuntu/Debian, OL, SLES:

     ```
     echo never > /sys/kernel/mm/transparent_hugepage/defrag
     echo never > /sys/kernel/mm/transparent_hugepage/enabled
     ```

Modify the permissions of the rc.local file:

```
chmod +x /etc/rc.d/rc.local
```

2. If your cluster hosts are using the GRUB bootloader, modify the GRUB configuration to disable THP:
   a) Add the following line to the GRUB_CMDLINE_LINUX options in the /etc/default/grub file:

   ```
   transparent_hugepage=never
   ```

   b) Run the following command:

   ```
   grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

3. You can also disable transparent hugepages interactively (but remember this will not survive a reboot).

   To disable transparent hugepages temporarily as root:

   ```
   # echo 'never' > DEFRAG_FILE_PATHNAME
   # echo 'never' > ENABLED_FILE_PATHNAME
   ```

   To disable transparent hugepages temporarily using sudo:

   ```
   $ sudo sh -c "echo 'never' > DEFRAG_FILE_PATHNAME"
   $ sudo sh -c "echo 'never' > ENABLED_FILE_PATHNAME"
   ```

Setting the vm.swappiness Linux Kernel Parameter

The Linux kernel parameter, vm.swappiness, is a value from 0-100 that controls the swapping of application data (as anonymous pages) from physical memory to virtual memory on disk. You can set the value of the vm.swappiness parameter for minimum swapping.

The higher the parameter value, the more aggressively inactive processes are swapped out from physical memory. The lower the value, the less they are swapped, forcing filesystem buffers to be emptied.

On most systems, vm.swappiness is set to 60 by default. This is not suitable for Hadoop clusters because processes are sometimes swapped even when enough memory is available. This can cause lengthy garbage collection pauses for important system daemons, affecting stability and performance.

Cloudera recommends that you set vm.swappiness to a value between 1 and 10, preferably 1, for minimum swapping on systems where the RHEL kernel is 2.6.32-642.el6 or higher.

To view your current setting for vm.swappiness, run:

```
cat /proc/sys/vm/swappiness
```

To set vm.swappiness to 1, run:

```
sudo sysctl -w vm.swappiness=1
```

To ensure persistence of the vm.swappiness value after reboot:

```
echo 'vm.swappiness=1' > /etc/sysctl.d/90-cloudera-swappiness.conf
```

Swap space allocation

Cloudera recommends following the guidelines provided by your operating system vendor to configure the swap space on each host. If your vendor recommends a swap space range, then use the lowest recommended value.



Note: If your operating system vendor does not have a recommendation for swap space, then search online for *"Linux default swap space"* and see the recommendation provided by Red Hat.

File system partitioning recommendations

This section helps you to understand the recommendations and set up the file system partitions on master and worker nodes on a CDP Private Cloud Base cluster.

Partitioning recommendations for all nodes

- Root partition: OS and core program files
- Swap: Size 2X system memory

Partitioning recommendations for worker nodes

- Hadoop worker node: Hadoop must have its partitions for Hadoop files and logs. Drives must be partitioned using XFS, ext4, or ext3 in that order of preference.

  Note: You must not use LVM because it adds latency.

- Worker nodes: All Hadoop partitions must be mounted individually from drives in the /grid/[0-n] format.

Hadoop worker node partitioning configuration example

- /swap: Cloudera recommends following the guidelines provided by your operating system vendor to configure the swap space on each host. If your vendor recommends a swap space range, then use the lowest recommended value.

  Note: If your operating system vendor does not have a recommendation for swap space, then search online for "Linux default swap space" and see the recommendation provided by Red Hat.

- /root: 20 GB (sufficient space for existing files, future log file growth, and OS upgrades)
- /grid/0/: [full disk GB] first partition for Hadoop to use for local storage
- /grid/1/: second partition for Hadoop to use
- /grid/2/: third partition for Hadoop to use, and so on

Redundancy (RAID) recommendations

- Master nodes: Configured for reliability (RAID 10, dual Ethernet cards, dual power supplies, and so on.)
- Worker nodes: RAID is not necessary as the cluster manages the worker nodes' failure automatically. Data is stored across at least three different hosts, therefore redundancy is built-in. Worker nodes must be built for speed and low cost.

Related Information

Selecting the underlying Linux file system
CentOS
Hadoop Reference Architectures

Improving Performance in Shuffle Handler and IFile Reader

The MapReduce shuffle handler and IFile reader use native Linux calls, (posix_fadvise(2) and sync_data_range), on Linux systems with Hadoop native libraries installed.

Shuffle Handler

You can improve MapReduce shuffle handler performance by enabling shuffle readahead. This causes the TaskTracker or Node Manager to pre-fetch map output before sending it over the socket to the reducer.

- To enable this feature for YARN, set mapreduce.shuffle.manage.os.cache, to true (default). To further tune performance, adjust the value of mapreduce.shuffle.readahead.bytes. The default value is 4 MB.
- To enable this feature for MapReduce, set the mapred.tasktracker.shuffle.fadvise to true (default). To further tune performance, adjust the value of mapred.tasktracker.shuffle.readahead.bytes. The default value is 4 MB.

IFile Reader

Enabling IFile readahead increases the performance of merge operations. To enable this feature for either MRv1 or YARN, set mapreduce.ifile.readahead to true (default). To further tune the performance, adjust the value of mapreduce.ifile.readahead.bytes. The default value is 4MB.

Tips and Best Practices for Jobs

This section describes changes you can make at the job level.

Use the Distributed Cache to Transfer the Job JAR

Use the distributed cache to transfer the job JAR rather than using the JobConf(Class) constructor and the JobConf.setJar() and JobConf.setJarByClass() methods.

To add JARs to the classpath, use -libjars *JAR1,JAR2*. This copies the local JAR files to HDFS and uses the distributed cache mechanism to ensure they are available on the task nodes and added to the task classpath.

The advantage of this, over JobConf.setJar, is that if the JAR is on a task node, it does not need to be copied again if a second task from the same job runs on that node, though it will still need to be copied from the launch machine to HDFS.



Note: -libjars works only if your MapReduce driver uses ToolRunner. If it does not, you would need to use the DistributedCache APIs (Cloudera does not recommend this).

For more information, see item 1 in the blog post *How to Include Third-Party Libraries in Your MapReduce Job*.

Changing the Logging Level on a Job (MRv1)

You can change the logging level for an individual job. You do this by setting the following properties in the job configuration:

- mapreduce.map.log.level
- mapreduce.reduce.log.level

Valid values are NONE, INFO, WARN, DEBUG, TRACE, and ALL.

Example:

```
JobConf conf = new JobConf();
...

conf.set("mapreduce.map.log.level", "DEBUG");
conf.set("mapreduce.reduce.log.level", "TRACE");
...
```

Decrease Reserve Space

By default, the ext3 and ext4 filesystems reserve 5% space for use by the root user. This reserved space counts as Non DFS Used.

To view the reserved space use the tune2fs command:

```
# tune2fs -l /dev/sde1 | egrep "Block size:|Reserved block count"
Reserved block count:  36628312
Block size:            4096
```

The Reserved block count is the number of ext3/ext4 filesystem blocks that are reserved. The block size is the size in bytes. In this example, 150 GB (139.72 Gigabytes) are reserved on this filesystem.

Cloudera recommends reducing the root user block reservation from 5% to 1% for the DataNode volumes. To set reserved space to 1% with the tune2fs command:

```
# tune2fs -m 1 /dev/sde1
```

Choosing and Configuring Data Compression

Guidelines for compression types.

Guidelines for Choosing Data Compression

- GZIP compression uses more CPU resources than Snappy or LZO, but provides a higher compression ratio. GZip is often a good choice for cold data, which is accessed infrequently. Snappy or LZO are a better choice for hot data, which is accessed frequently.
- BZip2 can also produce more compression than GZip for some types of files, at the cost of some speed when compressing and decompressing. HBase does not support BZip2 compression.
- Snappy often performs better than LZO. It is worth running tests to see if you detect a significant difference.

Related Information

Hadoop File Formats Support

Managing Cloudera Manager

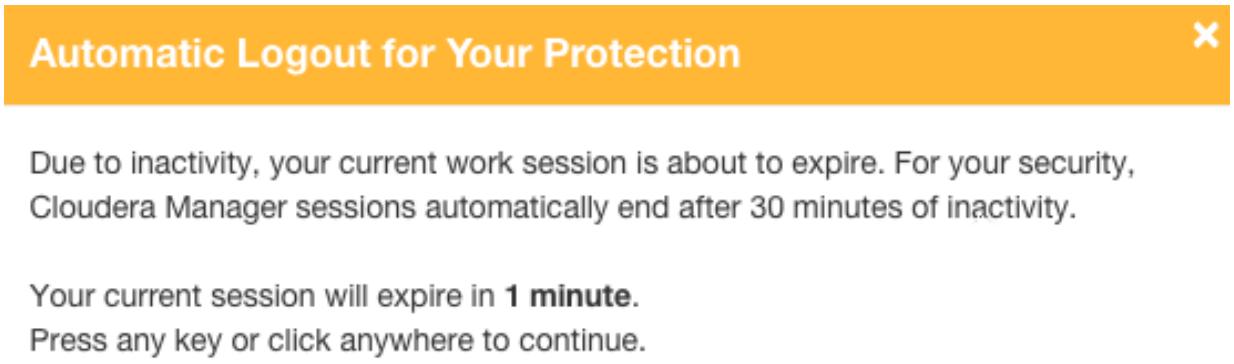
Automatic Logout

For security purposes, Cloudera Manager automatically logs out a user session after 30 minutes. You can change this session logout period.

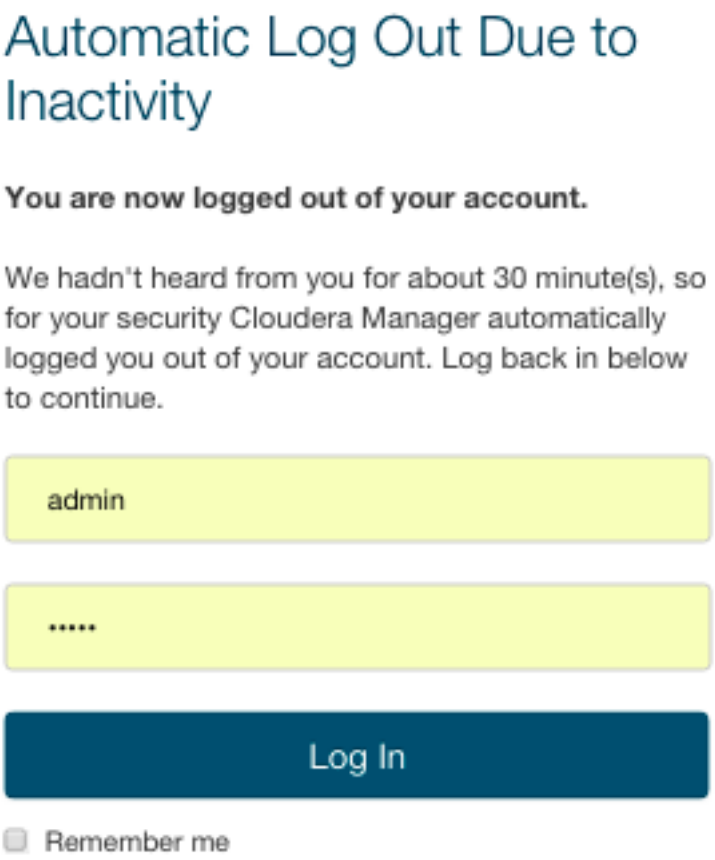
Procedure

1. Click AdministrationSettings.
2. Click CategorySecurity.
3. Edit the Session Timeout property.
4. Enter a Reason for change, and then click Save Changes to commit the changes.

   When the timeout is one minute from triggering, the user sees the following message:



   If the user does not click the mouse or press a key, the user is logged out of the session and the following message appears:



Starting, Stopping, and Restarting the Cloudera Manager Server

To start the Cloudera Manager Server:

```
sudo systemctl start cloudera-scm-server
```

You can stop (for example, to perform maintenance on its host) or restart the Cloudera Manager Server without affecting the other services running on your cluster. Statistics data used by activity monitoring and service monitoring will continue to be collected during the time the server is down.

To stop the Cloudera Manager Server:

```
sudo systemctl stop cloudera-scm-server
```

To restart the Cloudera Manager Server:

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

Configuring Cloudera Manager

From the Administration menu you can select options for configuring settings that affect how Cloudera Manager interacts with your clusters.

Settings

The Settings page provides a number of categories as follows:

- Performance - Set the Cloudera Manager Agent heartbeat interval.
- Advanced - Enable API debugging and other advanced options.
- Monitoring - Set Agent health status parameters. For configuration instructions, see the topic *Configuring Cloudera Manager Agents*.
- Security - Set TLS encryption settings to enable TLS encryption between the Cloudera Manager Server, Agents, and clients. For configuration instructions, see Configuring TLS Encryption for Cloudera Manager Using Auto-TLS. You can also:
  - Set the realm for Kerberos security and point to a custom keytab retrieval script.
  - Specify session timeout and a "Remember Me" option.
- Ports and Addresses - Set ports for the Cloudera Manager Admin Console and Server. For configuration instructions, see Configuring Cloudera Manager Server Ports on page 79.
- Other
  - Enable Cloudera usage data collection For configuration instructions, see *Managing Anonymous Usage Data Collection*.
  - Set a custom header color and banner text for the Admin console.
  - Set an "Information Assurance Policy" statement – this statement will be presented to every user before they are allowed to access the login dialog box. The user must click "I Agree" in order to proceed to the login dialog box.
  - Disable/enable the auto-search for the Events panel at the bottom of a page.
- Support
  - Configure diagnostic data collection properties. See *Diagnostic Data Collection*.
  - Configure how to access Cloudera Manager help documentation.
- External Authentication - Specify the configuration to use LDAP, Active Directory, or an external program for authentication.
- Parcels - Configure settings for parcels, including the location of remote repositories that should be made available for download, and other settings such as the frequency with which Cloudera Manager will check for new parcels, limits on the number of downloads or concurrent distribution uploads. See Overview of Parcels on page 102 for more information.
- Network - Configure proxy server settings. See Configuring Network Settings for a Proxy Server on page 79.
- Custom Service Descriptors - Configure custom service descriptor properties for Cloudera Manager Add-on Services.

You can also configure the following:

- Alerts
- Users
- Kerberos
- License

  See Managing Licenses and Cluster Registration on page 111.
- Language

  You can change the language of the Cloudera Manager Admin Console User Interface through the language preference in your browser. Information on how to do this for the browsers supported by Cloudera Manager is shown under the Administration page. You can also change the language for the information provided with activity and health events, and for alert email messages by selecting Language, selecting the language you want from the drop-down list on this page, then clicking Save Changes.

Related Information

Configuring Cloudera Manager Agents

Managing Anonymous Usage Data Collection
Diagnostic Data Collection
Alerts

Configuring Cloudera Manager Server Ports

You can specify the ports used to access the Cloudera Manager Server using the Admin Console. You can also specify the port used by agents to connect to the Server.

About this task

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. Select AdministrationSettings.
2. Under the Ports and Addresses category, set the following options as described below:

Setting	Description
HTTP Port for Admin Console	Specify the HTTP port to use to access the Server using the Admin Console.
HTTPS Port for Admin Console	Specify the HTTPS port to use to access the Server using the Admin Console.
Agent Port to connect to Server	Specify the port for Agents to use to connect to the Server.

3. Click Save Changes.
4. Restart the Cloudera Manager Server.

Configuring Network Settings for a Proxy Server

How to configure a poxy server for connections to Cloudera Manager.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

To configure a proxy server thorough which data is downloaded to and uploaded from the Cloudera Manager Server, do the following:

1. Select AdministrationSettings.
2. Click the Network category.
3. Configure proxy properties.
4. Enter a Reason for change, and then click Save Changes to commit the changes.

Moving the Cloudera Manager Server to a New Host

You can move the Cloudera Manager Server if either the Cloudera Manager database server or a current backup of the Cloudera Manager database is available.

Procedure

1. Identify a new host on which to install Cloudera Manager.
2. Install Cloudera Manager on a new host, using the method described in the topic *Install the Cloudera Manager Server Packages*.



Important:

- The Cloudera Manager version on the destination host must match the version on the source host.
- Do not install the other components, such as CDH and databases.

3. Copy the entire contents of the /var/lib/cloudera-scm-server/ directory on the old host to that same path on the new host. Ensure you preserve permissions and all file contents.
4. Copy the entire contents of the local parcel directory on the old host to that same path on the new host. Ensure you preserve permissions and all file contents. (The default location is /opt/cloudera/parcel-repo but this can be configured with the Local Parcel Repository Path configuration property, under AdministrationSettings.)
5. If the database server is not available:
   a) Install the database packages on the host that will host the restored database. This could be the same host on which you have just installed Cloudera Manager or it could be a different host. If you used the embedded PostgreSQL database, install the PostgreSQL package as described in the topic *Managing the Embedded PostgreSQL Database*. If you used an external MySQL, PostgreSQL, or Oracle database, reinstall the database following the instructions in *Step 4: Install and Configure Databases*.
   b) Restore the backed up databases to the new database installation.
6. Update /etc/cloudera-scm-server/db.properties with the database name, database instance name, username, and password.
7. Do the following on all cluster hosts:
   a) In /etc/cloudera-scm-agent/config.ini, update the server_host property to the new hostname.
   b) If you are replacing the Cloudera Manager database with a new database, and you are not using a backup of the original Cloudera Manager database, delete the /var/lib/cloudera-scm-agent/cm_guid file.
   c) Restart the agent using the following command:

   ```
   sudo service cloudera-scm-agent restart
   ```
8. Stop the Cloudera Manager server on the source host by running the following command:

   ```
   service cloudera-scm-server stop
   ```
9. Copy any Custom Service Descriptor files for add-on services to the configured directory on the new Cloudera Manager host. The directory path is configured by going to AdministrationSettings and editing the Local Descriptor Repository Path property. The default value is/opt/cloudera/csd. See Add-on Services.
10. Start the Cloudera Manager Server on the new (destination) host. Cloudera Manager should resume functioning as it did before the failure. Because you restored the database from the backup, the server should accept the running state of the Agents, meaning it will not terminate any running processes.

    The process is similar with secure clusters, though files in /etc/cloudera-scm-server must be restored in addition to the database. See the *Security* documentation.
11. Set the new host name for the new Cloudera Manager Server by navigating to Administration Settings Server SSL Certificate Host Name and restart the Cloudera Manager Server.

Migrating Cloudera Manager Management Services to Another Host

Relocating and migrating data from Cloudera Manager Management Services to another host.

About this task

To move the existing Cloudera Manager Management Services to another host.

Before you begin

You must have the minimum requirements to proceed with the migration process.

- Cloudera Manager 5 and higher
- Database credentials are required for the configured RDBMS databases.



Note: CDH cluster services are not affected during this process.

Password information

Embedded database installations create passwords on the disk that can be collected in /etc/cloudera-scm-server/db.mgmt.properties. You must have these credentials to reapply the database configuration during the role move.

Directory locations

Default Management data directory locations are used in all commands, adjust your location as required for the process.



Important: Consider JVM Heap, data-dir storage requirements, and existing usage when determining the new Host node.

You can read more about how to determine these values here:

- Data Storage for Monitoring Data
- Reports Manager
- Event Server

These documents contain information about the data storage types and requirements discussed in this process.

You should understand them well so that you can properly size a new host for the Cloudera Management Service processes and data.

Procedure

1. Record the existing service database credentials like:
   - hostname
   - type
   - dbuser
   - dbpass
   - configuration values (heap and safety values)
2. From Cloudera Manager, navigate to > Clusters > Cloudera Management Services > Configuration> Reports Manager > Database

   Note: Record the following for a future step:
   - Type
   - Hostname
   - Database name
   - Username and password
3. From Cloudera Manager navigate to Clusters > Cloudera Management Services > Configuration > Alert Publisher

   Record the values for the following:
   - Mail Server Hostname
   - Username
   - Password
   - Recipients

   Confirm if the Java Keystore (If TLS is enabled).
4. From Cloudera Manager, navigate to Clusters > Cloudera Management Services > Cloudera Management Service (Service-Wide) > Security > SSL Client Truststore File Location.

   Attention: Record the location and jks file for the next step.
5. Copy Truststore file to the new HostB node in the same directory location, owner as

```
cloudera-scm:cloudera-scm

[HostA]# cd /var/lib/cloudera-scm-agent/agent-cert
```

```
[HostA]# tar -czpf /tmp/agent-cert.tar.gz agent-cert

[HostA]# scp /tmp/agent-cert.tar.gz HostB.example.com:/tmp

[HostB]# cd /var/lib/cloudera-scm-agent/agent-cert

[HostB]# tar -xf /tmp/agent-cert.tar.gz
```

6. Stop all the Cloudera Manager Management services: Cloudera Manager > Clusters > Cloudera Management Services > Actions > Stop
7. Relocate the data directories to the new host with the same permissions and owner.

```
cloudera-scm:cloudera-scm
```

Issue the commands on HostA to tar up preserving ownership and perms, later securely copy (SCP) the service data to Host B - (you might need to SCP the files between servers using another method.)

```
[HostA]# cd /var/lib

[HostA]# tar -czpf /tmp/cloudera-host-monitor-data-dir.tar.gz cloudera-hos
t-monitor

[HostA]# tar -czpf /tmp/cloudera-service-monitor-data-dir.tar.gz cloudera-
service-monitor

[HostA]# tar -czpf /tmp/cloudera-scm-eventserver-data-dir.tar.gz cloudera-
scm-eventserver
[HostA]# tar -czpf /tmp/cloudera-scm-navigator-data-dir.tar.gz cloudera-
scm-navigator

[HostA]# scp /tmp/*-data-dir.tar.gz HostB.example.com:/tmp
```

8. Issue the commands on Host B to uncompress the tarball into /var/lib on the new host.

```
[HostB]# cd /var/lib

[HostB]# tar -xf /tmp/*-data-dir.tar.gz
```

9. Remove the roles: Cloudera Manager > Clusters > Cloudera Management Services > Instances
   Select all > Actions for Selected > Delete



Note: Ignore validation errors that the services are missing.

10. Click Add Role instances.
11. Select the specific host under each service to assign the roles to nodes and click Next to continue.

    Mandatory Roles

    Some of the roles that are required for you to proceed further with the migration process.

    - Event Server
    - Alert Publisher
    - Host Monitor
    - Service Monitor
    - Reports Manager

12. Input and test any of the database credential collected from Step #1.

    For example, for the embedded database on HostA, the configuration for Rman would be as follows:

```
Database Host Name: HostA.example.com:7432
```

```
Database Type: PostgreSQL

Database Name : rman
Username: rman
Password: rman
```

13. Review each service and apply any remaining configuration specifics like Heap or safety valves noted earlier.
14. Click Finish to save the various changes.
15. Start all the Cloudera Manager Management services



Note: New on-disk data directories, log locations, log files and permissions for each role will be created at first run with empty schema's and will immediately start collecting data if the tarball archives are not already in place.

16. Confirm health and functionality of each service.

    Wait five minutes or more and the graphs should have data and the cluster status should be visible. Allow Service monitor time to run canary tests, some errors require several sequential positive results.

17. Use Cloudera Manager to refresh the pages and the graph data should reappear.
18. Cleanup and remove the old data and log directories from old hosts (HostA).

    Default Locations of Cloudera Management Service logs and data directories

    Index of Cloudera Management Service default locations for logs and data directories:

    - Event Server Index Directory (/var/lib/cloudera-scm-eventserver)
    - Host Monitor Storage Directory (/var/lib/cloudera-host-monitor)
    - Service Monitor Storage Directory (/var/lib/cloudera-service-monitor)
    - Alert Publisher has no on disk data, only log location and alert config required during Add Role Wizard
    - Reports Manager Working Directory (/var/lib/cloudera-scm-headlamp)
    - Reports Manager data in RDBMS, reconfiguration required during Add Role Wizard (working directory is temporary location for hourly processing, no data is stored there)
    - Alert Publisher Log Directory (/var/log/cloudera-scm-alertpublisher)
    - Event Server Log Directory (/var/log/cloudera-scm-eventserver)
    - Host Monitor Log Directory (/var/log/cloudera-scm-firehose)
    - Audit Log Directory (/var/log/cloudera-scm-navigator/audit)
    - Reports Manager Log Directory (/var/log/cloudera-scm-headlamp)
    - Service Monitor Log Directory (/var/log/cloudera-scm-firehose)

    Steps to change the passwords for a management service in embedded Postgres.

```
# export PGPASSWORD=$(head -1 /var/lib/cloudera-scm-server-db/data/gener
ated_password.txt)

# psql -U cloudera-scm -h localhost -p 7432 scm
scm=# ALTER USER rman with password 'rman';

ALTER ROLE

scm=# \q
```

Migrating from the Cloudera Manager Embedded PostgreSQL Database Server to an External PostgreSQL Database

Cloudera Manager provides an embedded PostgreSQL database server for demonstration and proof of concept deployments when creating a cluster. To remind users that this embedded database is not suitable for production,

Cloudera Manager displays the banner text: "You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database before moving into production."

If, however, you have already used the embedded database, and you are unable to redeploy a fresh cluster, then you must migrate to an external PostgreSQL database.

Note: This procedure does not describe how to migrate to a database server other than PostgreSQL. Moving databases from one database server to a different type of database server is a complex process that requires modification of the schema and matching the data in the database tables to the new schema. It is strongly recommended that you engage with Cloudera Professional Services if you wish to perform a migration to an external database server other than PostgreSQL.

Prerequisites

Before migrating the Cloudera Manager embedded PostgreSQL database to an external PostgreSQL database, ensure that your setup meets the following conditions:

- The external PostgreSQL database server is running.
- The database server is configured to accept remote connections.
- The database server is configured to accept user logins using md5.
- No one has manually created any databases in the external database server for roles that will be migrated.

  Note: To view a list of databases in the external database server (requires default superuser permission):

  ```
  sudo -u postgres psql -l
  ```

- All health issues with your cluster have been resolved.

For details about configuring the database server, see the topic *Configuring and Starting the PostgreSQL Server*.

Important: Only perform the steps in *Configuring and Starting the PostgreSQL Server*. Do not proceed with the creation of databases as described in the subsequent section.

For large clusters, Cloudera recommends running your database server on a dedicated host. Engage Cloudera Professional Services or a certified database administrator to correctly tune your external database server.

Step 1: Identify Roles that Use the Embedded Database Server

Before you can migrate to another database server, you must first identify the databases using the embedded database server. When the Cloudera Manager Embedded Database server is initialized, it creates the Cloudera Manager database and databases for roles in the Management Services. The Installation Wizard (which runs automatically the first time you log in to Cloudera Manager) or Add Service action for a cluster creates additional databases for roles when run. It is in this context that you identify which roles are used in the embedded database server.

Procedure

1. Obtain and save the cloudera-scm superuser password from the embedded database server. You will need this password in subsequent steps:

   ```
   head -1  /var/lib/cloudera-scm-server-db/data/generated_password.txt
   ```

2. Make a list of all services that are using the embedded database server. Then, after determining which services are not using the embedded database server, remove those services from the list. The scm database must remain in your list. Use the following table as a guide:

Table 2: Cloudera Manager Embedded Database Server Databases

Service	Role	Default Database Name	Default Username
Cloudera Manager Server		scm	scm

Service	Role	Default Database Name	Default Username
Cloudera Management Service	Activity Monitor	amon	amon
Hive	Hive Metastore Server	hive	hive
Hue	Hue Server	hue	7uu7uu7uhue
Cloudera Management Service	Navigator Audit Server	nav	nav
Cloudera Management Service	Navigator Metadata Server	navms	navms
Oozie	Oozie Server	oozie_oozie_server	oozie_oozie_server
Cloudera Management Service	Reports Manager	rman	rman
Sentry	Sentry Server	sentry	sentry

3. Verify which roles are using the embedded database. Roles using the embedded database server always use port 7432 (the default port for the embedded database) on the Cloudera Manager Server host.

   For Cloudera Management Services:

   a. Select Cloudera Management Service > Configuration, and type "7432" in the Search field.
   b. Confirm that the hostname for the services being used is the same hostname used by the Cloudera Manager Server.

   Note:

   If any of the following fields contain the value "7432", then the service is using the embedded database:

   - Activity Monitor
   - Navigator Audit Server
   - Navigator Metadata Server
   - Reports Manager

   For the Oozie Service:

   a. Select Oozie service > Configuration, and type "7432" in the Search field.
   b. Confirm that the hostname is the Cloudera Manager Server.

   For Hive, Hue, and Sentry Services:

   a. Select the specific service > Configuration, and type "database host" in the Search field.
   b. Confirm that the hostname is the Cloudera Manager Server.
   c. In the Search field, type “database port” and confirm that the port is 7432.
   d. Repeat these steps for each of the services (Hive, Hue and Sentry).

4. Verify the database names in the embedded database server match the database names on your list (Step 2). Databases that exist on the database server and not used by their roles do not need to be migrated. This step is to confirm that your list is correct.

   Note: Do not add the postgres, template0, or template1 databases to your list. These are used only by the PostgreSQL server.

```
psql -h localhost -p 7432 -U cloudera-scm -l
```

```
Password for user cloudera-scm: <password>
                                                List of databases
       Name        |       Owner        | Encoding |  Collate   |   Ctype
    |         Access
--------------------+--------------------+----------+------------+-------
-----+---------------
amon                | amon               | UTF8     | en_US.UTF8 | en_US.U
TF8 |
```

```
hive                | hive                | UTF8     | en_US.UTF8 | en_US.UT
F8 |
hue                 | hue                 | UTF8     | en_US.UTF8 | en_US
.UTF8 |
nav                 | nav                 | UTF8     | en_US.UTF8 | en_US.
UTF8 |
navms               | navms               | UTF8     | en_US.UTF8 | en_US.U
TF8 |
oozie_oozie_server | oozie_oozie_server | UTF8     | en_US.UTF8 | en_US.UT
F8 |
postgres            | cloudera-scm        | UTF8     | en_US.UTF8 | en_US
.UTF8 |
rman                | rman                | UTF8     | en_US.UTF8 | en_US.
UTF8 |
scm                 | scm                 | UTF8     | en_US.UTF8 | en_US.U
TF8 |
sentry              | sentry              | UTF8     | en_US.UTF8 | en_US.UT
F8 |
template0           | cloudera-scm        | UTF8     | en_US.UTF8 | en_US
.UTF8 | =c/"cloudera-scm"
template1           | cloudera-scm        | UTF8     | en_US.UTF8 | en_US.UT
F8 | =c/"cloudera-scm"
(12 rows)
```

Results

You should now have a list of all roles and database names that use the embedded database server, and are ready to proceed with the migration of databases from the embedded database server to the external PostgreSQL database server.

What to do next

Proceed to Step 2: Migrate Databases from the Embedded Database Server to the External PostgreSQL Database Server.

Step 2: Migrate Databases from the Embedded Database Server to the External PostgreSQL Database Server

After you identify the roles that use the embedded database server, you can migrate from the Cloudera Manager embedded database server to the external PostgreSQL database server. When you migrate, you export the PostgreSQL user roles from the embedded database, import the PostgreSQL user roles into the external database, import the Cloudera Manager database on the external database server, and perform other tasks.

About this task

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

While performing this procedure, ensure that the Cloudera Manager Agents remain running on all hosts. Unless otherwise specified, when prompted for a password use the cloudera-scm password.



Note: After completing this migration, you cannot delete the cloudera-scm postgres superuser unless you remove the access privileges for the migrated databases. Minimally, you should change the cloudera-scm pos tgres superuser password.

Procedure

1. In Cloudera Manager, stop the cluster services identified in the previous step as using the embedded database server. Be sure to stop the Cloudera Management Service as well. Also be sure to stop any services with dependencies on these services. The remaining CDH services will continue to run without downtime.



Note: If you do not stop the services from within Cloudera Manager before stopping Cloudera Manager Server from the command line, they will continue to run and maintain a network connection to the embedded database server. If this occurs, then the embedded database server will ignore any command line stop commands (Step 2) and require that you manually stop the process, which in turn causes the services to crash instead of stopping cleanly.

2. Navigate to Hosts > All Hosts, and make note of the number of roles assigned to hosts. Also take note whether or not they are in a commissioned state. You will need this information later to validate that your scm database was migrated correctly.
3. Stop the Cloudera Manager Server. To stop the server:

```
sudo service cloudera-scm-server stop
```

4. Obtain and save the embedded database superuser password (you will need this password in subsequent steps) from the generated_password.txt file:

```
head -1  /var/lib/cloudera-scm-server-db/data/generated_password.txt
```

5. Export the PostgreSQL user roles from the embedded database server to ensure the correct users, permissions, and passwords are preserved for database access. Passwords are exported as an md5sum and are not visible in plain text. To export the database user roles (you will need the cloudera-scm user password):

```
pg_dumpall -h localhost -p 7432 -U cloudera-scm -v --roles-only -f "/var/
tmp/cloudera_user_roles.sql"
```



Important:

If you see the version mismatch between the existing PostgreSQL Database server and the utilities of the new external PostgreSQL Database server, then Cloudera recommends creating a symlink with the related PostgreSQL utilities such as pg_dumpall, pg_dump, and pg_restore to ensure the `pg utility` commands work successfully.

Run the following command to create a symlink with the pg_dumpall utility:

```
sudo ln -sf /usr/pgsql-<existing PostgreSQL Database server version>/
bin/pg_dumpall /usr/bin/pg_dumpall
```

6. Edit /var/tmp/cloudera_user_roles.sql to remove any CREATE ROLE and ALTER ROLE commands for databases not in your list. Leave the entries for cloudera-scm untouched, because this user role is used during the database import.
7. Export the data from each of the databases on your list you created when you identified roles that use the embedded database server:

```
pg_dump -F c -h localhost -p 7432 -U cloudera-scm [database_name]  > /var/
tmp/[database_name]_db_backup-$(date +”%m-%d-%Y”).dump
```

The following is a sample data export command for the scm database:

```
pg_dump -F c -h localhost -p 7432 -U cloudera-scm scm > /var/tmp/scm_db_
backup-$(date +%m-%d-%Y).dump
```

```
Password:
```



Important:

If you see the version mismatch between the existing PostgreSQL Database server and the utilities of the new external PostgreSQL Database server, then Cloudera recommends creating a symlink with the related PostgreSQL utilities such as pg_dumpall, pg_dump, and pg_restore to ensure the `pg utility` commands work successfully.

Run the following command to create a symlink with the pg_dump utility:

```
sudo ln -sf /usr/pgsql-<existing PostgreSQL Database server version>/
bin/pg_dump /usr/bin/pg_dump
```

8. Stop and disable the embedded database server:

```
service cloudera-scm-server-db stop
chkconfig cloudera-scm-server-db off
```

Confirm that the embedded database server is stopped:

```
netstat -at | grep 7432
```

9. Back up the Cloudera Manager Server database configuration file:

```
cp /etc/cloudera-scm-server/db.properties /etc/cloudera-scm-server/db.pr
operties.embedded
```

10. Copy the file /var/tmp/cloudera_user_roles.sql and the database dump files from the embedded database server host to /var/tmp on the external database server host:

```
cd /var/tmp
scp cloudera_user_roles.sql *.dump <user>@<postgres-server>:/var/tmp
```

11. Import the PostgreSQL user roles into the external database server.

The external PostgreSQL database server superuser password is required to import the user roles. If the superuser role has been changed, you will be prompted for the username and password.



Note: Only run the command that applies to your context; do not execute both commands.

- To import users when using the default PostgreSQL superuser role:

```
sudo -u postgres psql -f /var/tmp/cloudera_user_roles.sql
```

- To import users when the superuser role has been changed:

```
psql -h <database-hostname> -p <database-port> -U <superuser> -f /var/tm
p/cloudera_user_roles.sql
```

For example:

```
psql -h pg-server.example.com -p 5432 -U postgres -f /var/tmp/cloudera_u
ser_roles.sql
```

```
Password for user postgres
```

12. Import the Cloudera Manager database on the external server. First copy the database dump files from the Cloudera Manager Server host to your external PostgreSQL database server, and then import the database data:

Note: To successfully run the pg_restore command, there must be an existing database on the database server to complete the connection; the existing database will not be modified. If the -d <existing-database> option is not included, then the pg_restore command will fail.

```
pg_restore -C -h <database-hostname> -p <database-port> -d <existing-database> -U cloudera-scm -v <data-file>
```

Repeat this import for each database.

The following example is for the scm database:

```
pg_restore -C -h pg-server.example.com -p 5432 -d postgres -U cloudera-scm -v /var/tmp/scm_server_db_backup-20180312.dump
```

```
pg_restore: connecting to database for restore
Password:
```

Important:

If you see the version mismatch between the existing PostgreSQL Database server and the utilities of the new external PostgreSQL Database server, then Cloudera recommends creating a symlink with the related PostgreSQL utilities such as pg_dumpall, pg_dump, and pg_restore to ensure the `pg utility` commands work successfully.

Run the following command to create a symlink with the pg_restore utility:

```
sudo ln -sf /usr/pgsql-<existing PostgreSQL Database server version>/bin/pg_restore /usr/bin/pg_restore
```

13. Update the Cloudera Manager Server database configuration file to use the external database server. Edit the /etc/cloudera-scm-server/db.properties file as follows:
    a) Update the com.cloudera.cmf.db.host value with the hostname and port number of the external database server.
    b) Change the com.cloudera.cmf.db.setupType value from "EMBEDDED" to "EXTERNAL".
14. Start the Cloudera Manager Server and confirm it is working:

```
service cloudera-scm-server start
```

Note that if you start the Cloudera Manager GUI at this point, it may take up to five minutes after executing the start command before it becomes available.

In Cloudera Manager Server, navigate to Hosts > All Hosts and confirm the number of roles assigned to hosts (this number should match what you found in Step 2); also confirm that they are in a commissioned state that matches what you observed in Step 2.

15. Update the role configurations to use the external database hostname and port number. Only perform this task for services where the database has been migrated.

- For Cloudera Management Services:
  a. Select Cloudera Management Service > Configuration, and type "7432" in the Search field.
  b. Change any database hostname properties from the embedded database to the external database hostname and port number.
  c. Click Save Changes.
- For the Oozie Service:
  a. Select Oozie service > Configuration, and type "7432" in the Search field.
  b. Change any database hostname properties from the embedded database to the external database hostname and port number.
  c. Click Save Changes.

- For Hive, Hue, and Sentry Services:
  a. Select the specific service > Configuration, and type "database host" in the Search field.
  b. Change the hostname from the embedded database name to the external database hostname.
  c. Click Save Changes.

16. Start the Cloudera Management Service and confirm that all management services are up and no health tests are failing.
17. Start all Services via the Cloudera Manager web UI. This should start all services that were stopped for the database migration. Confirm that all services are up and no health tests are failing.
18. On the embedded database server host, remove the embedded PostgreSQL database server:
    a) Make a backup of the /var/lib/cloudera-scm-server-db/data directory:

```
tar czvf /var/tmp/embedded_db_data_backup-$(date +"%m-%d-%Y").tgz /var/lib/cloudera-scm-server-db/data
```

    b) Remove the embedded database package:

    For RHEL/SLES:

```
rpm --erase cloudera-manager-server-db-2
```

    For Debian/Ubuntu:

```
apt-get remove cloudera-manager-server-db-2
```

    c) Delete the /var/lib/cloudera-scm-server-db/data directory.

Migrating from the Cloudera Manager External PostgreSQL Database Server to a MySQL/Oracle Database Server

Cloudera Manager provides an embedded PostgreSQL database server for demonstration and proof of concept deployments when creating a cluster. To remind users that this embedded database is not suitable for production, Cloudera Manager displays the banner text: "You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database before moving into production."

If you have already used the embedded database, and you are unable to redeploy a fresh cluster, then you must migrate to an external PostgreSQL database.



Note: You can migrate to an external MySQL or Oracle database only after successfully migrating from the embedded PostgreSQL database server to the external PostgreSQL database server.

Prerequisites

Before migrating from the Cloudera Manager external PostgreSQL database to an external MySQL/Oracle database, ensure that your setup meets the following conditions:

- Configuration uses Cloudera Manager 5.15.0 or later on supported platforms.
- You must have a valid Cloudera Manager Enterprise license.
- If Cloudera Manager is secured, then you must import Kerberos account manager credentials and regenerate them.
- You must have a destination host installed with the supported database of choice (MySQL or Oracle). For details about installing and configuring MySQL for Cloudera, see the topic *Install and Configure MySQL for Cloudera Software*. For details about installing and configuring Oracle for Cloudera, see the topic *Install and Configure Oracle Database Software for Cloudera Software*.
- You have made configured target database hosts available.
- You have planned for cluster downtime during the migration process.

- You have a plan to follow service specific database migration instructions for services other than Cloudera Manager. Refer to the appropriate service migration documentation for your cluster setup.
- No one has manually created any databases in the external database server for roles that will be migrated.
- All health issues with your cluster are resolved.

For large clusters, Cloudera recommends running your database server on a dedicated host. Engage Cloudera Professional Services or a certified database administrator to correctly tune your external database server.

Migrate from the Cloudera Manager External PostgreSQL Database Server to a MySQL/Oracle Database Server

When you migrate from the Cloudera Manager external PostgreSQL database server to a MySQL or Oracle database server, you export the Cloudera Manager configuration, prepare the target database for Cloudera Manager, and complete other tasks.

About this task

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

Before you begin

1. Migrate from the embedded PostgreSQL database server to an external PostgreSQL database server as described in the topic *Migrating from the Cloudera Manager Embedded PostgreSQL Database Server to an External PostgreSQL Database*.

   Important: Migrating directly from the Cloudera Manager embedded PostgreSQL to a MySQL or Oracle database is not supported. You must first migrate from the Cloudera Manager embedded PostgreSQL database server to the external PostgreSQL database server. After performing this migration, you can use this procedure to migrate from the external PostgreSQL database server to MySQL or Oracle database servers.
2. Upgrade the Cloudera Manager enterprise license by navigating to Administration Licenses and installing a valid Cloudera Manager license.

Procedure

1. 

   Important:

   By default, Cloudera Manager redacts sensitive information from the exported configuration JSON file. If you need the ability to restore the Cloudera Manager configuration, you must do one of the following:

   - Disable redaction. The JSON file will contain all the configurations, including sensitive information and can be used to restore the Cloudera Manager configuration.
   - Replace the redacted information. The JSON will contain the word "REDACTED" where sensitive information was redacted. Replace these values with the correct values before restoring the Cloudera Manager configuration.

   Perform the following steps to disable redaction by setting a JVM parameter:

   a) Log in the Cloudera Manager server host using ssh.

   b) Edit the /etc/default/cloudera-scm-server file by adding the following property (separate each property with a space) to the line that begins with export CMF_JAVA_OPTS.

   ```
   -Dcom.cloudera.api.redaction=false
   ```

   For example:

   ```
   export CMF_JAVA_OPTS="-Xmx2G -Dcom.cloudera.api.redaction=false"
   ```

c) Restart Cloudera Manager Server:

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

2. Export your Cloudera Manager Configuration. First, get the latest supported API version:

```
curl -u <admin_username>:<admin_password> "http://<cm_server_host>:7180/
api/version"
```

```
curl -u <admin_username>:<admin_password> "http://<cm_server_host>:7180/
api/<api_version> /cm/deployment" > <path_to_file>/cm-deployment.json
```

The following is an example of the API version command:

```
curl -u admin:admin "http://10.17.103.191:7180/api/v19/cm/deployment" > /
root/cm-deployment.json
```



Important:

If you have Cloudera Manager with TLS for the Admin Console enabled, retrieve the certificate file and use curl with the --cacert option:

```
curl --cacert <certificate_file> -u admin:admin "https://<cm_server_
host>:7183/api/version"
```

3. Preserve Cloudera Manager's GUID by running the following command on the Cloudera Manager server to create a /etc/cloudera-scm-server/uuid file:

```
sudo -u postgres psql -qtAX scm -c "select GUID from CM_VERSION" > uuid
```



Important: Confirm the name of your Cloudera Manager database in /etc/cloudera-scm-server/db.properties.

Move the UUID file to Cloudera Manager server's /etc/cloudera-scm-server directory.

```
sudo mv uuid /etc/cloudera-scm-server/
```

4. Stop the cluster and the Cloudera Management services. For details, see *Starting, Stopping, Refreshing, and Restarting a Cluster* and *Stopping the Cloudera Management Service*.
5. Run the following command to stop the Cloudera Manager server:

```
sudo systemctl stop cloudera-scm-server
```

6. Prepare the target database for Cloudera Manager. For details, see *Install and Configure MySQL for Cloudera Software* or *Install and Configure Oracle Database for Cloudera Software*.
7. Run the following command to obtain the value of the GUID:

```
cat /etc/cloudera-scm-server/uuid
```

8. Insert the GUID returned from the previous command into MySQL or Oracle as follows:
   a) Run the following command to login to MySQL:

```
sudo -u mysql -p <name_of_cm_database>
```

   b) Run the following command to login to Oracle:

```
sqlplus system@localhost
Enter password: [***password***]
```

c) Run the following command to update the GUID:

```
update cm_version set guid = '[***GUID***]'
```

9. The process directory (/var/run/cloudera-scm-agent/process/) must be cleaned out for all hosts with agents running on them. The agent completes this cleanup with a server reboot. However, if a server reboot is not a viable option, do the following steps to accomplish the same task.
   a) Run the following command on all hosts to stop the agent and supervisor:

```
sudo systemctl stop cloudera-scm-agent
```

   b) Run the following command to confirm that the agent and supervisor processes are stopped:

```
ps -ef | grep -i cmf-agent; ps -ef | grep -i supervisor
```

   c) Move the existing /var/run/cloudera-scm-agent/ directory:

```
mv /var/run/cloudera-scm-agent /var/run/cloudera-scm-agent-BU
```

   The agent recreates the directory. Delete the backed-up copy after confirming that the migration was successful.
   d) Run the following command on all hosts to start the agent and supervisor:

```
sudo systemctl start cloudera-scm-agent
```

10. Run the following command to start the Cloudera Manager server:

```
sudo systemctl start cloudera-scm-server
```

11. Login to Cloudera Manager. Exit the installation wizard by clicking the product logo in the upper-left corner to stop the wizard and return to the Cloudera Manager home page.
12. Run the following command to restore the Cloudera Manager configuration:

```
curl -H "Content-Type: application/json" --upload-file <path_to_file>/cm
-deployment.json -u <admin_username>:<admin_password> "http://<cm_server
_host>:7180/api/<api_version>/cm/deployment?deleteCurrentDeployment=true"
```

    The following example shows how to restore a Cloudera Manager configuration:

```
curl -H "Content-Type: application/json" --upload-file /root/cm-deployme
nt.json -u admin:admin "http://172.31.113.146:7180/api/v19/cm/deployment?
deleteCurrentDeployment=true"
```

13. Start the following: Cloudera Management Service, Host Monitor, and Services Monitor. Verify that all the services in the Cloudera Management Service started and are Healthy. For details, see *Starting the Cloudera Management Service* and *Starting and Stopping Cloudera Management Service Roles*.
14. Go to Home Status and select the cluster(s) that you previously stopped, and select Start from the Actions drop-down menu.

Related Information

Migrating from the Cloudera Manager Embedded PostgreSQL Database Server to an External PostgreSQL Database
Starting, Stopping, Refreshing, and Restarting a Cluster
Stopping the Cloudera Management Service
Install and Configure MySQL for Cloudera Software
Install and Configure Oracle Database for Cloudera Software
Starting the Cloudera Management Service
Starting and Stopping Cloudera Management Service Roles

Managing Cloudera Manager Server Logs

You can use the Cloudera Manager Server logs to troubleshoot problems with Cloudera Manager .

Viewing the Cloudera Manager Server Logs

To help you troubleshoot problems, you can view the Cloudera Manager Server log. You can view the logs in the **Logs** page or in specific pages for the log.

1. In the left menu, click Diagnostics Logs .
2. Next to Sources, select the Cloudera Manager Server checkbox and deselect the other options.
3. Adjust the search criteria and click Search.

You can also view the raw Cloudera Manager Server log by logging in to the Cloudera Manager Server host and view the /var/log/cloudera-scm-server/cloudera-scm-server.log file.

Setting the Cloudera Manager Server Log Location

You can customize the location of the Cloudera Manager Server log files.

Depending on your requirements, choose one of the following methods:

- Log-Only Relocation (Recommended): Changes only the log directory path (for example, to /var/log/custom/cloudera-scm-server) without modifying operational state or run directories.
- Broad Directory Relocation: Changes the parent directory for both logs and operational run directories using the CMF_VAR variable.

Option 1: Relocating only the Cloudera Manager Server Logs (Recommended)

To change the log directory location without altering other operational paths, you can either move your existing log directory to the new location or create a new directory from scratch:
Method A: Moving the existing Log Directory (Preserves existing Logs)

1. Stop the Cloudera Manager Server:

```
sudo systemctl stop cloudera-scm-server
```

2. Move the existing log directory to your preferred target location:

```
sudo mv /var/log/cloudera-scm-server /var/log/custom/cloudera-scm-server
```

3. Open /etc/default/cloudera-scm-server in a text editor and configure the custom log path by appending the -Dcmf.log.dir system property to CMF_JAVA_OPTS:

```
export CMF_JAVA_OPTS="$CMF_JAVA_OPTS -Dcmf.log.dir=/var/log/custom/cloudera-scm-server"
```

4. Start the Cloudera Manager Server:

```
sudo systemctl start cloudera-scm-server
```

5. Verify that the Cloudera Manager Server log files are actively being written to your new custom directory:

```
ls -l /var/log/custom/cloudera-scm-server/cloudera-scm-server.log
```

Method B: Creating a new Custom Log Directory

1. Stop the Cloudera Manager Server:

```
sudo systemctl stop cloudera-scm-server
```

2. Create your custom log directory and set ownership to cloudera-scm:

```
sudo mkdir -p /var/log/custom/cloudera-scm-server
```

```
sudo chown -R cloudera-scm:cloudera-scm /var/log/custom/clou
dera-scm-server
```

3. Open /etc/default/cloudera-scm-server in a text editor and configure the custom log path by appending the -Dcmf.log.dir system property to CMF_JAVA_OPTS:

```
export CMF_JAVA_OPTS="$CMF_JAVA_OPTS -Dcmf.log.dir=/var/log/
custom/cloudera-scm-server"
```

4. Start the Cloudera Manager Server:

```
sudo systemctl start cloudera-scm-server
```

5. Verify that the Cloudera Manager Server log files are actively being written to your new custom directory:

```
ls -l /var/log/custom/cloudera-scm-server/cloudera-scm-serve
r.log
```

Option 2: Relocating the Parent Variable Directory (CMF_VAR)



Important: Setting CMF_VAR changes the base path for both log directories (/var/log/cloudera-scm-server) and run (state) directories (/var/run/cloudera-scm-server). It does not alter or move /var/lib/cloudera-scm-server (data or library storage remains untouched under /var/lib). Use this option only if you intend to relocate all /var-dependent log and runtime state directories.

1. Stop the Cloudera Manager Server:

```
sudo systemctl stop cloudera-scm-server
```

2. Move or copy the existing log and run directories to the new base location (for example, /opt) to retain state and existing log files, or create new directories and assign ownership:

```
sudo mkdir -p /opt/log/cloudera-scm-server
```

```
sudo mkdir -p /opt/run/cloudera-scm-server
```

```
sudo chown -R cloudera-scm:cloudera-scm /opt/log /opt/run
```

3. Open /etc/default/cloudera-scm-server in a text editor and set the CMF_VAR environment variable to the new parent directory:

```
export CMF_VAR=/opt
```

4. Start the Cloudera Manager Server:

```
sudo systemctl start cloudera-scm-server
```

Configuring Cloudera Manager Server log retention

You can customize the maximum file size and the number of backup log files retained for Cloudera Manager Server logs by modifying the /etc/cloudera-scm-server/log4j.properties configuration file.

Cloudera Manager Server log appenders

Cloudera Manager Server uses specific Log4j appenders to manage its log output files:

Log appender	Associated log file	Default MaxFileSize	Default MaxBackupIndex	Description
LOGFILE	cloudera-scm-server.log	10MB	10	Main Cloudera Manager Server log file.
JETTYLOG	cmf-server--nio.log	10MB	10	Embedded Jetty web server access and operational logs.
PERFLOGFILE	cmf-server-perf.log	10MB	10	Performance metrics and execution timing logs.
CG1	cloudera-scm-server-gc.log	10MB	1	Garbage Collection (GC) logging output.
MEMFILE	cmf-server-memory.log	10MB	10	Memory usage and heap tracking logs.

Modify log retention parameters

Perform the following steps to configure maximum file sizes and retention limits in the /etc/cloudera-scm-server/log4j.properties configuration file:

1. Log in to the Cloudera Manager Server host through SSH.
2. Open the /etc/cloudera-scm-server/log4j.properties configuration file in a text editor:

```
sudo vi /etc/cloudera-scm-server/log4j.properties
```

3. Locate the appender properties you want to modify and update the MaxFileSize (file size before rollover) and MaxBackupIndex (number of rotated log files to retain):

```
# Main Server Log (cloudera-scm-server.log)
log4j.appender.LOGFILE.MaxFileSize=20MB
log4j.appender.LOGFILE.MaxBackupIndex=15

# Jetty Web Server Log (cmf-server--nio.log)
log4j.appender.JETTYLOG.MaxFileSize=10MB
log4j.appender.JETTYLOG.MaxBackupIndex=10

# Performance Log (cmf-server-perf.log)
log4j.appender.PERFLOGFILE.MaxFileSize=10MB
log4j.appender.PERFLOGFILE.MaxBackupIndex=10
# Garbage Collection Log
log4j.appender.CG1.MaxFileSize=10MB
log4j.appender.CG1.MaxBackupIndex=2

# Memory Tracking Log (cmf-server-memory.log)
log4j.appender.MEMFILE.MaxFileSize=10MB
log4j.appender.MEMFILE.MaxBackupIndex=10
```

4. Save and close the file.
5. Restart the Cloudera Manager Server daemon to apply the changes:

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

Related Information

Logs

Cloudera Manager Agents

The Cloudera Manager Agent is a Cloudera Manager component that works with the Cloudera Manager Server to manage the processes that map to role instances.

In a Cloudera Manager managed cluster, you can only start or stop role instance processes using Cloudera Manager. Cloudera Manager uses a modified version of the open source process management tool, Supervisor. The Cloudera Manager Agent manages this via a process called cloudera-scm-supervisord, which starts processes, redirects log files, notifies of process failures, and sets the effective user ID for service processes. Cloudera Manager supports automatically restarting a crashed process. It will also flag a role instance with a bad health flag if its process crashes repeatedly right after start up.

The Agent is started by init.d at start-up. It, in turn, contacts the Cloudera Manager Server and determines which processes should be running. The Agent is monitored as part of Cloudera Manager's host monitoring. If the Agent stops heartbeating, the host is marked as having bad health.

One of the Agent's main responsibilities is to start and stop processes. When the Agent detects a new process from the Server heartbeat, the Agent creates a directory for it in /var/run/cloudera-scm-agent and unpacks the configuration. It then contacts cloudera-scm-supervisord, which starts the process.

Monitored File Systems

Cloudera Manager Agent monitors file systems to report usage and free space. The following is a list of the supported file system mount types:

- ext2
- ext3
- ext4
- xfs
- aufs
- tmpfs

By modifying the parameter monitored_nodev_filesystem_types in the Cloudera Manager Agent's config.ini file, you can also add NFS type mounts to the monitored list. For more information on this parameter, see Configuring Cloudera Manager Agents.

Other file system types, especially auto-mounted file systems, or other network file systems, are not supported and might prevent the operation of the Cloudera Manager Agent.

cm_processes

To enable Cloudera Manager to run scripts in subdirectories of /var/run/cloudera-scm-agent, (because /var/run is mounted noexec in many Linux distributions), Cloudera Manager mounts a tmpfs, named cm_processes, for process subdirectories.

A tmpfs defaults to a max size of 50% of physical RAM but this space is not allocated until its used, and tmpfs is paged out to swap if there is memory pressure.

The lifecycle actions of cmprocesses can be described by the following statements:

- Created when the Agent starts up for the first time with a new cloudera-scm-supervisord process.
- If it already exists without noexec, reused when the Agent is started using start and not recreated.
- Remounted if Agent is started using clean_restart.
- Unmounting and remounting cleans out the contents (since it is mounted as a tmpfs).
- Unmounted when the host is rebooted.

- Not unmounted when the Agent is stopped.

Related Information

tmpfs

Starting, Stopping, and Restarting Cloudera Manager Agents

Starting Agents

To start Agents, the cloudera-scm-supervisord process, and all managed service processes, use the following command:

- Start

```
sudo systemctl start cloudera-scm-agent
```

Stopping and Restarting Agents

To stop or restart Agents while leaving the managed processes running, use one of the following commands:

- Stop

```
sudo systemctl stop cloudera-scm-agent
```

- Restart

```
sudo systemctl restart cloudera-scm-agent
```

Hard Stopping and Restarting Agents



Warning: The hard_stop and hard_restart commands stop all running managed service processes on the host(s) where the command is run.

To stop or restart Agents, the cloudera-scm-supervisord process, and all managed service processes, use one of the following commands:

- Hard Stop
  RHEL 7, SLES 12, Ubuntu 18.04 and higher

```
sudo systemctl stop cloudera-scm-supervisord.service
sudo systemctl stop cloudera-scm-agent
```

- Hard Restart
  RHEL 7, SLES 12, Ubuntu 18.04 and higher

```
sudo systemctl stop cloudera-scm-supervisord.service
sudo systemctl restart cloudera-scm-agent
```



Important: Starting the Agent automatically launches cloudera-scm-supervisord.

Hard restart is useful for the following situations:

- You are upgrading Cloudera Manager and the cloudera-scm-supervisord code has changed between your current version and the new one. To properly complete this upgrade, you must restart the cloudera-scm-supervisord process as well.
- The cloudera-scm-supervisord freezes and needs to be restarted.

- You want to clear out all running state pertaining to Cloudera Manager and managed services.

Checking Agent Status

To check the status of the Agent process, use the command:

```
sudo systemctl status cloudera-scm-agent
```

Configuring Cloudera Manager Agents

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Cloudera Manager Agents can be configured globally using properties you set in the Cloudera Manager Admin Console and by setting properties in Agent configuration files.

Configuring Agent Heartbeat and Health Status Options

You can configure the Cloudera Manager Agent heartbeat interval and timeouts to trigger changes in Agent health as follows:

1. Select AdministrationSettings.
2. Under the Performance category, set the following option:

Property	Description
Send Agent Heartbeat Every	The interval in seconds between each heartbeat that is sent from Cloudera Manager Agents to the Cloudera Manager Server. Default: 15 sec.

3. Under the Monitoring category, set the following options:

Property	Description
Set health status to Concerning if the Agent heartbeats fail	The number of missed consecutive heartbeats after which a Concerning health status is assigned to that Agent. Default: 5.
Set health status to Bad if the Agent heartbeats fail	The number of missed consecutive heartbeats after which a Bad health status is assigned to that Agent. Default: 10.

4. Click Save Changes.

Configuring the Host Parcel Directory

Important: If you modify the parcel directory location, make sure that all hosts use the same location. Using different locations on different hosts can cause unexpected problems.

To configure the location of distributed parcels:

1. Click Hosts in the top navigation bar.
2. Click the Configuration tab.
3. Select CategoryParcels.
4. Configure the value of the Parcel Directory property. The setting of the parcel_dir property in the Cloudera Manager Agent configuration file overrides this setting (see below).
5. Enter a Reason for change, and then click Save Changes to commit the changes.
6. Restart the Cloudera Manager Agent on all hosts.

Agent Configuration File

The Cloudera Manager Agent supports different types of configuration options in the /etc/cloudera-scm-agent/config.ini file. You must update the configuration on each host. After changing a property, restart the Agent:

```
sudo systemctl restart cloudera-scm-agent
```

Section	Property	Description
[General]	server_host, server_port, listening_port, listening_hostname, listening_ip	Hostname and ports of the Cloudera Manager Server and Agent and IP address of the Agent. Also see Configuring Cloudera Manager Server Ports and Ports used by Cloudera Manager. The Cloudera Manager Agent configures its hostname automatically. You can also manually specify the hostname the Cloudera Manager Agent uses by updating the listening_hostname property. To manually specify the IP address the Cloudera Manager Agent uses, update the listening_ip property in the same file. To have a CNAME used throughout instead of the regular hostname, an Agent can be configured to use listening_hostname=*CNAME*. In this case, the CNAME should resolve to the same IP address as the IP address of the hostname on that machine. Users doing this will find that the host inspector will report problems, but the CNAME will be used in all configurations where that's appropriate. This practice is particularly useful for users who would like clients to use namenode.*MYCLUSTER.COMPANY*.com instead of machine1234.*MYCLUSTER.COMPANY*.com. In this case, namenode.*MYCLUSTER* would be a CNAME for machine1234.*MYCLUSTER*, and the generated client configurations (and internal configurations as well) would use the CNAME.
	lib_dir	Directory to store Cloudera Manager Agent state that persists across instances of the agent process and system reboots. The Agent UUID is stored here. Default: /var/lib/cloudera-scm-agent.
	local_filesystem_whitelist	The list of local filesystems that should always be monitored. Default: ext2,ext3,ext4.
	log_file	The path to the Agent log file. If the Agent is being started using the init.d script, /var/log/cloudera-scm-agent/cloudera-scm-agent.out will also have a small amount of output (from before logging is initialized). Default: /var/log/cloudera-scm-agent/cloudera-scm-agent.log.
	max_collection_wait_seconds	Maximum time to wait for all metric collectors to finish collecting data. Default: 10 sec.
	metrics_url_timeout_seconds	Maximum time to wait when connecting to a local role's web server to fetch metrics. Default: 30 sec.
	parcel_dir	Directory to store unpacked parcels. Default: /opt/cloudera/parcels. This property overrides the setting in Cloudera Manager. To use the recommended procedure, you must make sure that this property is commented out in each host config.ini file.
	supervisord_port	The port used by the cloudera-scm-supervisord process. A change to this value takes effect only after a hard restart of the cloudera-scm-supervisord process; a standard Agent restart is insufficient to apply this change. Default: 19001.
	task_metrics_timeout_seconds	Maximum time to wait when connecting to a local TaskTracker to fetch task attempt data. Default: 5 sec.

Section	Property	Description
	monitored_nodev_filesystem_types	List of non-device file system types to monitor. When hosts have NFS mounts, Cloudera Manager Agent might get stuck querying information about the NFS mounts. In this case, reduce the list of file systems in this parameter to only tmpfs. If you want to monitor NFS file systems, then change the list to tmpfs,nfs,nfs4.
[Security]	use_tls,verify_cert_file, client_key_file, client_keypw_file, client_cert_file	Security-related configuration. See • Configuring TLS Encryption for Cloudera Manager Using Auto-TLS • Adding a Host to a Cluster on page 35
[Cloudera]	mgmt_home	Directory to store Cloudera Management Service files. Default: /usr/share/cmf.
[JDBC]	cloudera_mysql_connector_jar, cloudera_oracle_connector_jar, cloudera_postgresql_jdbc_jar	Location of JDBC drivers. Default: • MySQL - /usr/share/java/mysql-connector-java.jar • Oracle - /usr/share/java/oracle-connector-java.jar • PostgreSQL - /usr/share/cmf/lib/postgresql-*VERSION-BUILD*.jdbc4.jar

Related Information

Health Tests

Starting, Stopping, and Restarting Cloudera Manager Agents

CNAME

Managing the Cloudera Manager Agent Logs

To help you troubleshoot problems, you can view the Cloudera Manager Agent logs. You can view the logs in the Logs page or in specific pages for the logs.

Related Information

Configuring Cloudera Manager Agents

Viewing the Cloudera Manager Agent Logs

Use the procedure to view and search the logs from all Cloudera Manager agents managed by this instance of Cloudera Manager.

Procedure

1. In the left menu, click DiagnosticsLogs.
2. Click Select Sources to display the log source list.
3. Uncheck the All Sources checkbox.
4. Click ▸ to the left of Cloudera Manager and select the Agent checkbox.
5. Click Search.

What to do next

You can also view the Cloudera Manager Agent log at /var/log/cloudera-scm-agent/cloudera-scm-agent.log on the Agent hosts.

Setting the Cloudera Manager Agent Log Location

By default. the Cloudera Manager Agent log is stored in /var/log/cloudera-scm-agent/. If there is not enough space in that directory, you can change the location of the log file.

Procedure

1. Set the log_file property in the Cloudera Manager Agent configuration file:

```
log_file=/opt/log/cloudera-scm-agent/cloudera-scm-agent.log
```

2. Create log/cloudera-scm_agent directories and set the owner and group to cloudera-scm. For example, if the log is stored in /opt/log/cloudera-scm-agent, do the following:

```
sudo su
cd /opt
mkdir log
chown cloudera-scm:cloudera-scm log
mkdir /opt/log/cloudera-scm-agent
chown cloudera-scm:cloudera-scm log/cloudera-scm-agent
```

3. Restart the Agent:

```
sudo service cloudera-scm-agent restart
```

Overview of Parcels

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

A *parcel* is a binary distribution format containing the program files, along with additional metadata used by Cloudera Manager. The important differences between parcels and packages are:

- Parcels are self-contained and installed in a versioned directory, which means that multiple versions of a given parcel can be installed side-by-side. You can then designate one of these installed versions as the active one. With packages, only one package can be installed at a time so there is no distinction between what is installed and what is active.
- Parcels are required for rolling upgrades.
- You can install parcels at any location in the filesystem. They are installed by default in /opt/cloudera/parcels. In contrast, packages are installed in /usr/lib.
- When you install from the Parcels page, Cloudera Manager automatically downloads, distributes, and activates the correct parcel for the operating system running on each host in the cluster. All hosts that make up a logical cluster must run on the same major OS release to be covered by Cloudera Support. Cloudera Manager must run on the same major OS release as at least one of the clusters it manages, to be covered by Cloudera Support. The risk of issues caused by running different minor OS releases is considered lower than the risk of running different major OS releases. Cloudera recommends running the same minor release cross-cluster, because it simplifies issue tracking and supportability.



Important: Cloudera Manager manages parcels without the need for users to manipulate parcels in the filesystem. You might cause failures or unexpected behaviors in your cluster if you perform any of the following unsupported actions:

- Installing parcels within custom RPM packages and saving them to the Cloudera Manager parcel directory.
- Downloading parcels and manually placing them in the Cloudera Manager parcel directory.
- Manually adding, modifying, or deleting files within the root parcels directory or its subdirectories.

Parcels are available for CDH, Cloudera Runtime and for other managed services.

Advantages of Parcels

Because of their unique properties, parcels offer the following advantages over packages:

- Distribution of Cloudera Runtime as a single object - Instead of having a separate package for each component of Cloudera Runtime, parcels are distributed as a single object. This makes it easier to distribute software to a cluster that is not connected to the Internet.
- Internal consistency - All Cloudera Runtime components are matched, eliminating the possibility of installing components from different versions.
- Installation outside of /usr - In some environments, Hadoop administrators do not have privileges to install system packages. With parcels, administrators can install to /opt, or anywhere else.

  Note: With parcels, the path to the Cloudera Runtime libraries is /opt/cloudera/parcels/CDH/lib instead of the usual /usr/lib. Do not link /usr/lib/ elements to parcel-deployed paths, because the links can cause scripts that distinguish between the two paths to not work.
- Installation of Cloudera Runtime without sudo - Parcel installation is handled by the Cloudera Manager Agent running as root, so you can install Cloudera Runtime without sudo.
- Decoupled distribution from activation - With side-by-side install capabilities, you can stage a new version of Cloudera Runtime across the cluster before switching to it. This allows the most time-consuming part of an upgrade to be done ahead of time without affecting cluster operations, thereby reducing downtime.
- Rolling upgrades - Using packages requires you to shut down the old process, upgrade the package, and then start the new process. Errors can be difficult to recover from, and upgrading requires extensive integration with the package management system to function seamlessly. With parcels, when a new version is staged side-by-side, you can switch to a new minor version by simply changing which version of Cloudera Runtime is used when restarting each process. You can then perform upgrades with rolling restarts, in which service roles are restarted in the correct order to switch to the new version with minimal service interruption. Your cluster can continue to run on the existing installed components while you stage a new version across your cluster, without impacting your current operations. Major version upgrades (for example, CDH 5 to CDH 6) require full service restarts because of substantial changes between the versions. Finally, you can upgrade individual parcels or multiple parcels at the same time.
- Upgrade management - Cloudera Manager manages all the steps in a CDH or Cloudera Runtime version upgrade. With packages, Cloudera Manager only helps with initial installation.
- Additional components - Parcels are not limited to Cloudera Runtime. Add-on service parcels are also available.
- Compatibility with other distribution tools - Cloudera Manager works with other tools you use for download and distribution, such as Puppet. Or, you can download the parcel to Cloudera Manager Server manually if your cluster has no Internet connectivity and then have Cloudera Manager distribute the parcel to the cluster.

Parcel Life Cycle

To enable upgrades and additions with minimal disruption, parcels have following phases:

- Downloaded -The parcel software is copied to a local parcel directory on the Cloudera Manager Server, where it is available for distribution to other hosts in any of the clusters managed by this Cloudera Manager Server. You can have multiple parcels for a product downloaded to your Cloudera Manager Server. After a parcel has been downloaded to the Server, it is available for distribution on all clusters managed by the Server. A downloaded parcel appears in the cluster-specific section for every cluster managed by this Cloudera Manager Server.
- Distributed - The parcel is copied to the cluster hosts, and components of the parcel are unpacked. Distributing a parcel does not upgrade the components running on your cluster; the current services continue to run unchanged. You can have multiple parcels distributed on your cluster. Distributing parcels does not require Internet access; the Cloudera Manager Agent on each cluster member downloads the parcels from the local parcel repository on the Cloudera Manager Server.
- Activated - Links to the parcel components are created. Activation does not automatically stop the current services or perform a restart. You can restart services after activation, or the system administrator can determine when to perform those operations.
- In Use - The parcel components on the cluster hosts are in use when you start or restart the services that use those components.
- Deactivated - The links to the parcel components are removed from the cluster hosts.
- Removed - The parcel components are removed from the cluster hosts.
- Deleted - The parcel is deleted from the local parcel repository on the Cloudera Manager Server.

Cloudera Manager detects when new parcels are available. You can configure Cloudera Manager to download and distribute parcels automatically. .

Parcel Locations

The default location for the local parcel directory on the Cloudera Manager Server is /opt/cloudera/parcel-repo. To change this location, follow the instructions in Configuring Cloudera Manager Server Parcel Settings on page 109.

The default location for the distributed parcels on managed hosts is /opt/cloudera/parcels. To change this location, set the parcel_dir property in /etc/cloudera-scm-agent/config.ini file of the Cloudera Manager Agent and restart the Cloudera Manager Agent or by following the instructions in Configuring the Host Parcel Directory on page 110.

Managing Parcels

Procedures for managing Parcels.

On the Parcels page in Cloudera Manager, you can manage parcel installation and activation and determine which parcel versions are running across your clusters. The Parcels page displays a list of parcels managed by Cloudera Manager. Cloudera Manager displays the name, version, and status of each parcel and provides available actions on the parcel.

Cloudera Manager Agent employs the flood - BitTorrent file distribution daemon to distribute parcels.

When the number of hosts grows large, download of parcels takes place in a consistent manner. The BitTorrent file distribution helps avoid choking Cloudera Manager with large number of download requests from agents.

The BitTorrent file distribution attempts to fetch pieces from all registered torrents either from its peers or from the seed HTTP URL (if provided).

Accessing the Parcels Page

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

Access the Parcels page by doing one of the following:

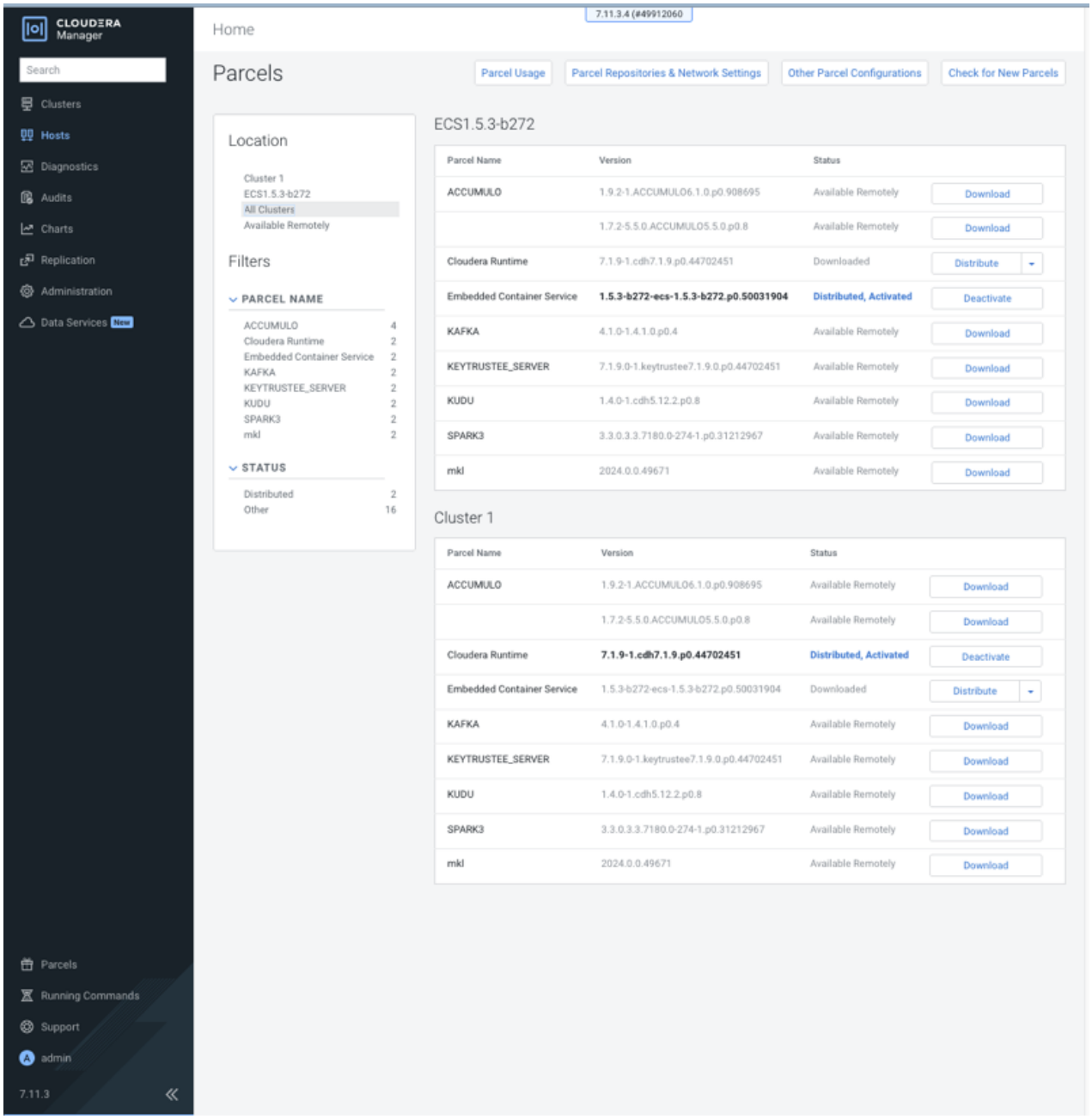
- Click the parcel icon in the top navigation bar.
- Click the Hosts in the top navigation bar, then the Parcels tab.

Use the selectors on the left side of the console to filter the displayed parcels:

- Location selector - View only parcels that are available remotely, only parcels pertaining to a particular cluster, or parcels pertaining to all clusters. When you access the Parcels page, the selector is set to Available Remotely.
- Error Status section of the Filters selector - Limit the list of displayed parcels by error status.
- Parcel Name section of the Filters selector - Limit the list of displayed parcels by parcel name.
- Status section of the Filters selector - Limit the list to parcels that have been distributed, parcels that have not been distributed (Other), or all parcels.

When you download a parcel, it appears in the list for each cluster managed by Cloudera Manager, indicating that the parcel is available for distribution on those clusters. Only one copy of the downloaded parcel resides on the Cloudera Manager Server. After you distribute the parcel, Cloudera Manager copies the parcel to the hosts in that cluster.

For example, if Cloudera Manager is managing two clusters, the rows in the All Clusters page list the information about the parcels on the two clusters. The Status column displays the current status of the parcels. The Version column displays version information about the parcel. Click the information icon to view the release notes for the parcel. The last column shows actions you can perform on the parcels, such as download, distribute, delete, deactivate, and remove from host.



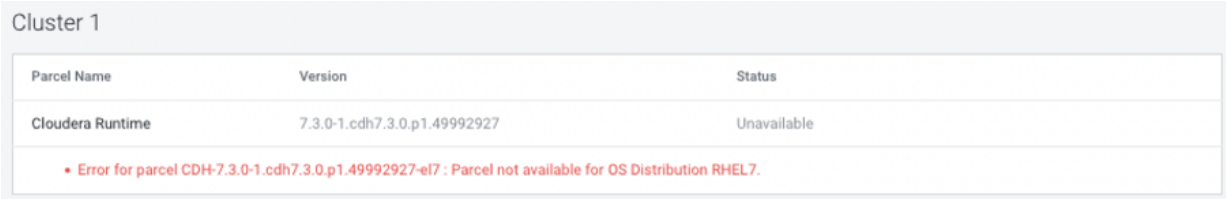
Downloading a Parcel

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

1. Go to the Parcels page. In the Location selector, click *CLUSTERNAME* or Available Remotely. Parcels that are available for download display the Available Remotely status and a Download button.

   If the parcel you want is not shown here—for example, you want to upgrade to a version of CDH that is not the most current version—you can make additional remote parcel repositories available. You can also configure the location of the local parcel repository and other settings. See Parcel Configuration Settings on page 109.

   If a parcel version is not compatible with the Cloudera Manager version, then the parcel appears with a red error message:



Such parcels are also listed when you select the Error status in the Error Status section of the Filters selector.

2. Click the Download button of the parcel you want to download to your local repository. The status changes to Downloading.

After a parcel has been downloaded, it is removed from the Available Remotely page.

Distributing a Parcel

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

Downloaded parcels can be distributed to the hosts in your cluster and made available for activation. Parcels are downloaded to the Cloudera Manager Server, so with multiple clusters, the downloaded parcels are shown as available to all clusters managed by the Cloudera Manager Server. However, you select distribution to a specific cluster's hosts on a cluster-by-cluster basis.

1. From the Parcels page, in the Location selector, select the cluster where you want to distribute the parcel, or select All Clusters. (The first cluster in the list is selected by default when you open the Parcels page.)
2. Click Distribute for the parcel you want to distribute. The status changes to Distributing. During distribution, you can:
   - Click the Details link in the Status column to view the Parcel Distribution Status page.
   - Click Cancel to cancel the distribution. When the Distribute action completes, the button changes to Activate, and you can click the Distributed status link to view the status page.

Distribution does not require Internet access; the Cloudera Manager Agent on each cluster member downloads the parcel from the local parcel repository hosted on the Cloudera Manager Server.

If you have a large number of hosts to which parcels must be distributed, you can control how many concurrent uploads Cloudera Manager performs. See Parcel Configuration Settings on page 109.

To delete a parcel that is ready to be distributed, click the triangle at the right end of the Distribute button and select Delete. This deletes the parcel from the local parcel repository.

Distributing parcels to the hosts in the cluster does not affect the current running services.

Activating a Parcel

Parcels that have been distributed to the hosts in a cluster are ready to be activated.

1. From the Parcels page, in the Location selector, choose *CLUSTERNAME* or All Clusters, and click the Activate button for the parcel you want to activate. This updates Cloudera Manager to point to the new software, which is ready to run the next time a service is restarted. A pop-up indicates which services must be restarted to use the new parcel.
2. Choose one of the following:
   - Restart - Activate the parcel and restart services affected by the new parcel.
   - Activate Only - Active the parcel. You can restart services at a time that is convenient. If you do not restart services as part of the activation process, you must restart them at a later time. Until you restart services, the current parcel continues to run.
3. Click OK.

Activating a new parcel also deactivates the previously active parcel for the product you just upgraded. However, until you restart the services, the previously active parcel displays a status of Still in use because the services are using that parcel, and you cannot remove the parcel until it is no longer being used.

If the parcel you activate updates the software for only a subset of services, even if you restart all of that subset, the previously active parcel displays Still in use until you restart the remaining services. For example, if you are running HDFS, YARN, Oozie, Hue, Impala, and Spark services, and you activate a parcel that updates only the Oozie service, the pop-up that displays instructs you to restart only the Oozie and Hue services. Because the older parcel is still in use by the HDFS, YARN, Impala, and Spark services, the parcel page shows that parcel as Still in use until you restart these remaining services.

Sometimes additional upgrade steps may be required. In this case, instead of Activate, the button will say Upgrade.

Deactivating a Parcel

You can deactivate an active parcel; this updates Cloudera Manager to point to the previous software version, which is ready to run the next time a service is restarted. From the Parcels page, choose *CLUSTERNAME* or All Clusters in the Location selector, and click the Deactivate button on an activated parcel.

To use the previous version of the software, restart your services.

Important: If you originally installed from parcels, and one version of the software is installed (that is, no packages, and no previous parcels have been activated and started), when you attempt to restart after deactivating the current version, your roles will be stopped and will not be able to restart.

Removing a Parcel

From the Parcels page, in the Location selector, choose *CLUSTERNAME* or All Clusters, click the to the right of an Activate button, and select Remove from Hosts.

Deleting a Parcel

From the Parcels page, in the Location selector, choose *CLUSTERNAME* or All Clusters, and click the to the right of a Distribute button, and select Delete.

Warning:

Note: Do not remove the Cloudera Runtime parcel entirely from Cloudera Manager unless it is no longer in use on any other cluster managed by this instance of Cloudera Manager. To remove a parcel from specific managed hosts, select the Remove from Hosts option instead of the Delete option.

Changing the Parcel Directory

The default location of the parcel directory is /opt/cloudera/parcels. To relocate distributed parcels to a different directory, do the following:

1. Stop all services in the cluster by performing the following steps:
   a. On the Cloudera Manager home page, navigate to the Status tab, click to the right of the cluster name and select Stop from the list view.
   b. Click Stop.
   c. Click Close.
2. Deactivate all in-use parcels.
3. Shut down the Cloudera Manager Agent on all hosts.
4. Move the existing parcels to the new location.
5. Configure the host parcel directory.
6. Start the Cloudera Manager Agents.
7. Activate the parcels.
8. Start all services in the cluster by performing the following steps:

a. On the Cloudera Manager home page, navigate to the Status tab, click ⋮ to the right of the cluster name and select Start from the list view.
b. Click Start.
c. Click Close.

Troubleshooting

If you experience an error while performing parcel operations, click the red 'X' icons on the parcel page to display a message that identifies the source of the error.

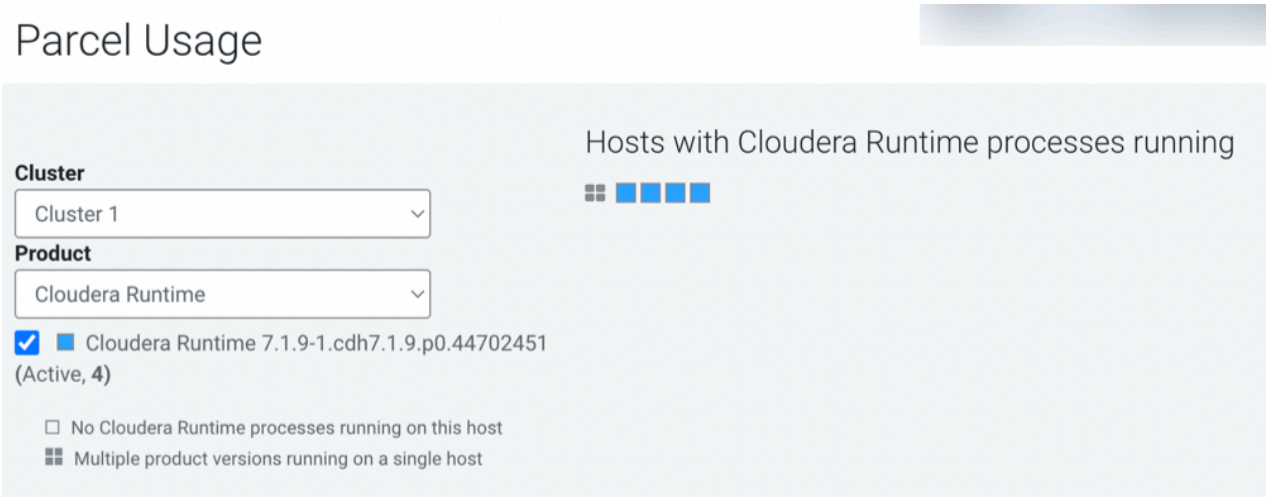
If a parcel is being distributed but never completes, make sure you have enough free space in the parcel download directories, because Cloudera Manager will try to download and unpack parcels even if there is insufficient space.

Viewing Parcel Usage

The Parcel Usage page shows parcels in current use in your clusters. In a large deployment, this makes it easier to keep track of different versions installed across the cluster, especially if some hosts were not available when you performed an installation or upgrade, or were added later. To display the Parcel Usage page:

1. Do one of the following:
   - Click the parcel icon in the top navigation bar.
   - Click Hosts in the top navigation bar and click the Parcels tab.
2. Click the Parcel Usage button.

This page only shows the usage of parcels, not components that were installed as packages. If you select a cluster running packages, the cluster is not displayed, and instead you see a message indicating the cluster is not running parcels.



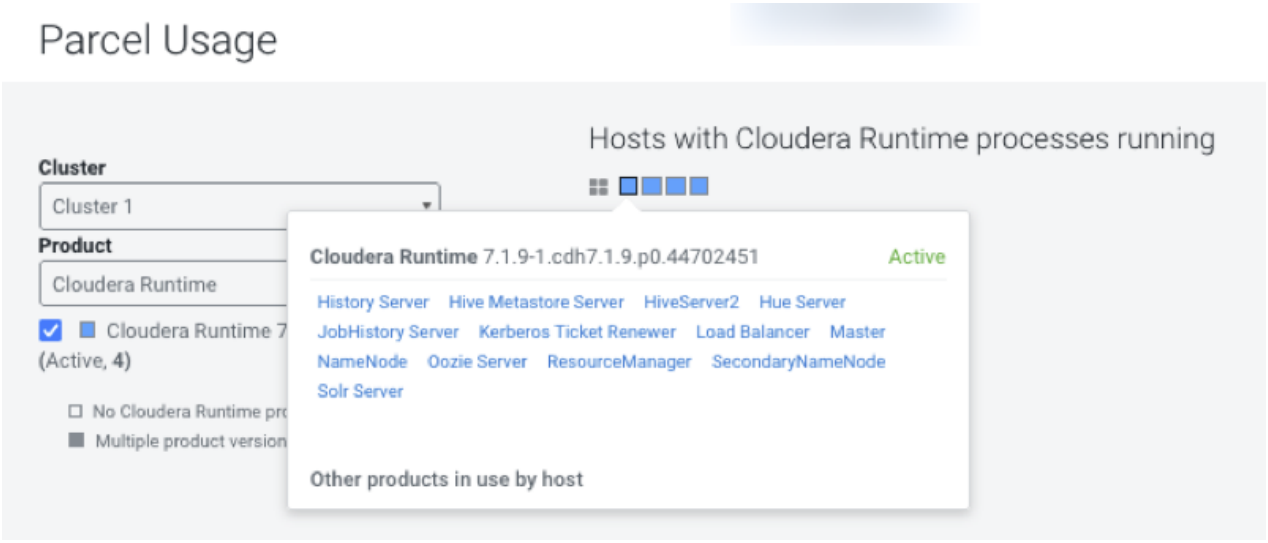
You can view parcel usage by cluster or by product.

You can also view just the hosts running only the active parcels, or just hosts running older parcels (not the currently active parcels), or both.

The host map at the right shows each host in the cluster, with the status of the parcels on that host. If the host is running the processes from the currently activated parcels, the host is indicated in blue. A black square indicates that a parcel has been activated, but that all the running processes are from an earlier version of the software. This occurs, for example, if you have not restarted a service or role after activating a new parcel. If you have individual hosts running components installed as packages, the square is empty.

Move the cursor over the grid icon to see the rack to which the hosts are assigned. Hosts on different racks are displayed in separate rows.

To view the exact versions of the software running on a given host, click the square representing the host. This displays the parcel versions installed on that host.



The pop-up lists the roles running on the selected host that are part of the listed parcel. Clicking a role opens the Cloudera Manager page for that role. It also shows whether the parcel is active or not.

Parcel Configuration Settings

You can configure where parcels are stored on the Cloudera Manager Server host, the URLs of parcel repositories, the properties of a proxy server through which parcels are downloaded, and where parcels distributed to cluster hosts are stored.

Configuring Cloudera Manager Server Parcel Settings

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

1. 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 Administration Settings .
     b. Select Category Parcels .
2. Specify a property:
   - Local Parcel Repository Path defines the path on the Cloudera Manager Server host where downloaded parcels are stored.
   - Remote Parcel Repository URLs is a list of repositories that Cloudera Manager checks for parcels. Initially this points to the latest released CDH 5 and CDH 6repositories, but you can add your own repository locations to the list. Use this mechanism to add Cloudera repositories that are not listed by default, such as older versions of CDH. You can also use this to add your own custom repositories. The locations of the Cloudera parcel repositories are https://archive.cloudera.com/*PRODUCT*/parcels/*VERSION* , where *PRODUCT* is a product name and *VERSION* is a specific product version, latest, or the substitution variable {latest_supported}. The substitution variable appears after the parcel for the CDH version with the same major number as the Cloudera Manager version to enable substitution of the latest supported maintenance version of CDH.

     To add a parcel repository:

     a. In the Remote Parcel Repository URLs list, click the addition symbol to open an additional row.

   b. Enter the path to the repository.
3. Click Save Changes.

You can also:

- Set the frequency with which Cloudera Manager checks for new parcels.
- Configure a proxy to access to the remote repositories.
- Configure whether downloads and distribution of parcels should occur automatically when new ones are detected. If automatic downloading and distribution are not enabled (the default), go to the Parcels page to initiate these actions.
- Control which products can be downloaded if automatic downloading is enabled.
- Control whether to retain downloaded parcels.
- Control whether to retain old parcel versions and how many parcel versions to retain

You can tune the parcel distribution load on your network by configuring the bandwidth limits and the number of concurrent uploads. The defaults are up to 50 MiB/s aggregate bandwidth and 50 concurrent parcel uploads.

- Theoretically, the concurrent upload count (Maximum Parcel Uploads) is unimportant if all hosts have the same speed Ethernet. Fifty concurrent uploads is acceptable in most cases. However, if the server has more bandwidth (for example, 10 GbE, and the normal hosts are using 1 GbE), then the count is important to maximize bandwidth. It should be at least the difference in speeds (10x in this case).
- The bandwidth limit (Parcel Distribution Rate Limit) should be your Ethernet speed (in MiB/seconds) divided by approximately 16. You can use a higher limit if you have QoS configured to prevent starving other services, or if you can accept the risk associated with higher bandwidth load.

Configuring a Proxy Server

To configure a proxy server thorough which data and parcels are downloaded to and uploaded from the Cloudera Manager Server, do the following:

1. Select AdministrationSettings.
2. Click the Network category.
3. Configure proxy properties.
4. Enter a Reason for change, and then click Save Changes to commit the changes.

Configuring the Host Parcel Directory



Important: If you modify the parcel directory location, make sure that all hosts use the same location. Using different locations on different hosts can cause unexpected problems.

To configure the location of distributed parcels:

1. Click Hosts in the top navigation bar.
2. Click the Configuration tab.
3. Select Category Parcels .
4. Configure the value of the Parcel Directory property. The setting of the parcel_dir property in the Cloudera Manager agent configuration file overrides this setting.
5. Enter a Reason for change, and then click Save Changes to commit the changes.
6. Restart the Cloudera Manager Agent on all hosts.

Configuring Peer-to-Peer Distribution of Parcels

Cloudera Manager uses a peer-to-peer service to efficiently distribute parcels to cluster hosts. The service is enabled by default and is configured to run on port 7191. You can change this port number, and you can disable peer-to-peer distribution.

To modify peer-to-peer distribution of parcels:

1. Open Cloudera Manager and select Hosts All Hosts Configuration .

2. Change the value of the P2P Parcel Distribution Port property to the new port number.

   Set the value to 0 to disable peer-to-peer distribution of parcels.
3. Enter a Reason for change, and then click Save Changes to commit the changes.

Configuring user and group creation for Parcels

You can configure how the Cloudera Manager agent manages permissions when reading Parcels. There are two configuration parameters available and both are enabled by default.

- Create Users and Groups for Parcels.

  This parameter determines whether Cloudera Manager creates users and groups for Parcels. This may not be desired if custom users and groups are being used, or if they are created externally.
- Apply Permissions with respect to files installed by the parcels

  This parameter determines whether Cloudera Manager applies permissions for files installed by the parcels.

If you are using custom users and groups, or using LDAP-based users and groups, de-select the Create Users and Groups for Parcels parameter and select the Apply Permissions with respect to files installed by the parcels parameter.



Note: Prior to Cloudera Manager version 7.7.1, there was only a single parameter, Create Users and Groups, and Apply File Permissions for Parcels.

Managing Licenses and Cluster Registration

When you install Cloudera Manager, you can either upload a valid Cloudera Base on premises license or select a free 60-day trial version.

Cloudera Base on premises supports the following license categories:

- Cloudera Base on premises Edition (V2 License): Starting in Cloudera Manager 7.13.2.10000 (Cloudera Runtime 7.3.2 SP1), Cloudera introduces V2 licenses, which require Mandatory Cluster Registration. Upon uploading a V2 license key or creating a new cluster, Cloudera Manager initiates a 60-day registration grace period.
  - Connected Environments: Cloudera Manager automatically completes registration with Cloudera servers.
  - Air-Gapped Environments: Administrators must manually register clusters in bulk using Cloudera Account 360.

  Important: If a cluster is not registered within the 60-day grace period, it enters a *Locked state*. In a *Locked state*, active data workloads continue to process data normally, but Cloudera Manager freezes all administrative management functions. You cannot start, restart, or configure cluster services until registration completes.
- Cloudera Base on premises Edition (Standard License): Legacy license format for earlier Cloudera Manager releases. When a standard license expires, you can no longer access the Cloudera Manager Admin Console to manage your clusters until you upload a valid license. However, your clusters continue to function, your data remains intact, and your running jobs keep executing.
- Cloudera Base on premises Edition Trial: A free 60-day trial that does not require an initial license file. When the 60-day trial period expires, access to the Cloudera Manager Admin Console is restricted until you upload a valid license. As with standard licenses, your underlying cluster workloads continue to run uninterrupted. You can use a trial license only once; after it expires or ends, you cannot restart the trial.

To purchase or renew a Cloudera Base on premises license, contact your Cloudera Account Team, fill out the *Contact Us* form on the Cloudera website, or call 866-809-4659.

Accessing the License Page

To access the license page, click AdministrationLicense.

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

If you have a license installed, the license page indicates its status (for example, whether your license is currently valid) and displays the license details: the license owner, the license key, the license start date, the expiration date, and the deactivation date. Typically the expiration date is the same as the deactivation date, at which point the Admin Console is no longer accessible. If the license expires, your clusters and data are unaffected.

Upgrading from a Cloudera Private Cloud Base Trial to Cloudera Private Cloud Base

You can upgrade your license from a trial license to a Cloudera Private Cloud Base license.

About this task

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

Procedure

1. Purchase a Cloudera Private Cloud Base Edition license from Cloudera.
2. On the License page, click Update License.
3. Click the Select License File field.
4. Browse to the location of your license file, click the file, and click Open.
5. Click Upload.

Renewing a License

You can upload a license file to renew a Cloudera Private Cloud Base license.

Before you begin

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

Procedure

1. Download the license file and save it locally.
2. In Cloudera Manager, go to the Home page.
3. Select AdministrationLicense.
4. Click Update License.
5. Browse to the license file you downloaded.
6. Click Upload.

Mandatory Cluster Registration Policy

Starting with Cloudera Manager 7.13.2 SP1 (7.13.2.10000) and Cloudera Runtime 7.3.2 SP1, Cloudera requires mandatory cluster registration when you apply a Version 2 (V2) license key. Cluster registration validates your environment against your software entitlements, links your deployments directly to your account, and optimizes license resource tracking across connected and air-gapped deployments.

Registration applies to all newly created clusters and existing clusters updating to a V2 license. Connected environments perform this process automatically within Cloudera Manager, whereas air-gapped environments require manual registration through Cloudera Account 360.

Understanding V2 Licenses and Admin Controls

When you apply a new V2 license key, Cloudera Manager introduces a Cluster Registration Status section and a dedicated Cluster Registration page in the Web Console.

Applying a V2 license key requires you to register:

- All newly created clusters.
- All existing, active clusters managed by that Cloudera Manager instance.

- All active clusters whenever you apply a renewed or updated V2 license key (Re-registration).

Understanding the 60-day Cluster Registration Grace Period

When you create a cluster or apply a V2 license key, Cloudera Manager initiates a *60-day registration grace period*.

Important: The 60-day grace period begins on the exact day you create the cluster or apply the license key in Cloudera Manager, not when Cloudera issues the license file.

Consequences of Non-Registration (Locked state)

If you do not register a cluster within the 60-day grace period, or if you apply a V2 license to an existing cluster after its 60-day period ends, Cloudera Manager places the cluster into a *Locked* state.

While a cluster remains in a *Locked* state, Cloudera Manager enforces the following operational restrictions:

- No service restarts: You cannot restart cluster services or roles.
- No service starts: You cannot start stopped services or clusters.
- Configuration freeze: You cannot save or deploy configuration changes across the cluster.
- GUI management lockout: Cloudera Manager locks administrative management functions in the Web Console.

Important: Running workloads continue to process data normally, but Cloudera Manager freezes all administrative control until you complete cluster registration.

Supported Registration Methods

Cloudera Manager supports two workflow paths based on your network architecture:

- Online Registration (Connected Environments): Cloudera Manager automatically connects directly to the Cloudera registration server to validate entitlements and complete registration in a single step (or through a single click on the Register button).
- Offline Registration (Air-Gapped Environments): For environments without direct internet access, you export a JSON package containing non-sensitive cluster identifiers from Cloudera Manager, process it through Cloudera Account 360 from an internet-connected workstation, and import the resulting confirmation package back into Cloudera Manager.

How to Renew and Deregister Clusters

To maintain accurate entitlement records and ensure proper resource usage, review the following key lifecycle terminology and procedures:

Important:
Registration Terminology:

- Re-register: Updating an existing, active cluster to associate it with a new or renewed license key (renewal).
- Deregister: Removing a cluster from Cloudera Account 360 to release license entitlements on the license server when decommissioning or deleting it.

License Renewals (Re-registering Active Clusters)

When you apply a renewed or updated V2 license key to Cloudera Manager, you must re-register all previously registered (active) clusters to associate them with the new key:

- Connected environments: Cloudera Manager automatically contacts Cloudera servers and re-registers all managed clusters.
- Air-Gapped environments: Administrators must perform an offline bulk registration process using Cloudera Account 360.

Deregistration and Decommissioning

To ensure accurate record-keeping and free up license entitlements for allocation to other clusters, you must explicitly deregister clusters when you decommission them:

- Deletion Requirement: The **Cluster Deregistration** page in Cloudera Manager only displays clusters that you have already deleted. Active clusters do not appear on this page.
- Resource Recovery: Completing the deregistration process releases the associated cluster entitlements on the Cloudera license server, making those nodes available for future cluster deployments.
- Air-Gapped Support: Like registration, cluster deregistration supports both automated online workflows for connected environments and manual offline workflows for air-gapped environments.
- Connected Environments: Deleting a cluster in Cloudera Manager automatically deregisters the cluster from Cloudera servers.
- Air-Gapped Environments: Administrators must manually select the deleted cluster on the **Cluster Deregistration** page, click Offline Deregistration, process the request package in Cloudera Account 360, and import the confirmation into Cloudera Manager.

Related Information

Registering and Deregistering Clusters in Cloudera Manager
FAQ for Cluster Registration and V2 Licenses

Registering and Deregistering Clusters in Cloudera Manager

Follow these procedures to register active clusters under a Version 2 (V2) license or deregister decommissioned clusters in Cloudera Manager.

Registering Clusters Online (Connected Environments)

Use this procedure to register one or more active clusters when your Cloudera Manager instance has direct outbound internet access to Cloudera servers.

Before you begin

- Role Requirement: You must hold Full Administrator privileges in Cloudera Manager.
- License File: You must upload a valid V2 license file to Cloudera Manager 7.13.2 SP1 (7.13.2.10000) or higher.
- Connectivity: Cloudera Manager must have outbound internet connectivity to Cloudera registration endpoints.

Steps

1. Log in to the Cloudera Manager Admin Console.
2. Navigate to the **Cluster Registration** page or locate the **Cluster Registration Status** section.
3. Select the checkboxes next to the specific clusters you want to register (or select all clusters).
4. Click Online Registration (or Register).
5. When the system prompts you with a confirmation dialog, click Confirm.

Cloudera Manager connects directly to the registration server, completes the registration process, updates the status in the Cluster Registration Status table to REGISTERED, and displays the timestamp of the last successful report.

Registering Clusters Offline in Air-Gapped Environments (Bulk Update)

Use this procedure to complete offline cluster registration or re-registration when you deploy a V2 license key in air-gapped Cloudera Manager environments. This workflow uses a two-way file exchange between your offline Cloudera Manager instance and an internet-connected browser session logged into Cloudera Account 360.

To avoid registering air-gapped clusters individually, Cloudera Manager provides the Bulk Cluster Key Migration feature. You can export a single JSON package containing details for all managed clusters, upload it once to Cloudera Account 360 using an internet-connected device, and import a single confirmation package back to register all clusters simultaneously.

Before you begin

- Role Requirement: You must hold Full Administrator privileges in Cloudera Manager.
- Account Access: You need valid login credentials for Cloudera Account 360 on an internet-connected device.
- License File: You must upload a valid V2 license file to Cloudera Manager 7.13.2 SP1 (7.13.2.10000) or higher.
- Transfer Media: A portable storage device or secure transfer mechanism to move files between your air-gapped network and the internet-connected workstation.

Steps

Follow these steps to complete the offline bulk registration process:

1. Generate and download the registration JSON package from your air-gapped Cloudera Manager environment.
   a. Log in to the air-gapped Cloudera Manager Admin Console.
   b. Navigate to the Cluster Registration page.
   c. Select the target clusters and click Offline Registration (or Bulk Migration).
   d. Click Download Registration Request.
   e. Save the generated .json file to your local machine or portable storage device.

   Important: This JSON package contains non-sensitive cluster identifiers that Cloudera Account 360 requires for entitlement matching. You can open and inspect the file prior to upload.
2. Upload the registration package to Cloudera Account 360 from an internet-connected device.
   a. Transfer the downloaded .json package to an internet-connected workstation.
   b. Open a web browser, log in to Cloudera Account 360, and navigate to License Management Bulk Cluster Key Migration (or the registration portal).
   c. Click Upload Registration Response and select the .json package you generated in Step 1 (or manually copy and paste the JSON content into the portal).
3. Validate and select your clusters for migration in Cloudera Account 360.
   a. 1. Review the parsed list of clusters that the system extracts from your JSON file.
      2. Select the specific clusters (or choose Select All) that you want to register under the new license key.
      3. Click Validate and Submit.

      Troubleshooting validation errors:

If any cluster in a bulk upload request fails entitlement validation, the system rejects the entire request. Cloudera Account 360 displays an explicit error message identifying the failing cluster. You must resolve the entitlement conflict or deselect the failing cluster before resubmitting the request.

4. Download the signed confirmation package from Cloudera Account 360.
   a. Once validation completes successfully, Cloudera Account 360 generates a signed registration confirmation file.
   b. Click Download Confirmation Package and save the file.
   c. Transfer the confirmation file back to your air-gapped Cloudera Manager environment.
5. Import the confirmation package into your air-gapped Cloudera Manager instance to complete registration.
   a. Return to the air-gapped Cloudera Manager Admin Console.
   b. On the **Offline Registration** prompt page, click Upload Confirmation.
   c. Select the confirmation package file you downloaded from Cloudera Account 360 and click Apply.
   d. Verify that Cloudera Manager validates the confirmation file.

Once you import the file, Cloudera Manager reads the confirmation, acknowledges the cluster registration with Cloudera, updates the status to REGISTERED, and unlocks the clusters to restore normal management functions.

Deregistering Deleted Clusters

Follow these steps to remove a decommissioned cluster from the Cloudera license server and release its associated license entitlements for reuse on other deployments.

Before you begin

- Role Requirement: You must hold Full Administrator privileges in Cloudera Manager.
- Deletion Requirement: You must delete the target cluster from Cloudera Manager before it appears on the deregistration page. Active clusters do not appear on this page.

Steps

1. In the Cloudera Manager Admin Console, delete the cluster you want to decommission.
2. Navigate to the **Cluster Deregistration** page.
3. Verify that the deleted cluster appears in the list.
4. Execute deregistration based on your environment type:
   - Connected Environments: Select the deleted cluster and click Deregister to automatically notify the license server and release entitlements.
   - Air-Gapped Environments: Select the deleted cluster, click Offline Deregistration, download the deregistration request JSON, upload and process it in Cloudera Account 360, and import the signed deregistration confirmation back into Cloudera Manager.

Once deregistered, the license server frees up the cluster node resources for allocation to other clusters.

Related Information

Mandatory Cluster Registration Policy
FAQ for Cluster Registration and V2 Licenses

Ending a Cloudera Private Cloud Base Trial

If you are using the trial edition, the License page indicates when your license will expire. However, you can end the trial at any time (prior to expiration) as follows:

About this task

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

Procedure

1. On the **License** page, click End Trial.
2. Confirm that you want to end the trial.
3. Restart the Cloudera Management Service, HBase, HDFS, and Hive services to pick up configuration changes.

FAQ for Cluster Registration and V2 Licenses

A collection of frequently asked questions about the mandatory cluster registration policy, the 60-day grace period, compliance enforcement, air-gapped workflows, and cluster deregistration introduced in Cloudera Manager 7.13.2 SP1 (7.13.2.10000) and Cloudera Base on premises 7.3.2 SP1.

What is Cluster Registration, and why does Cloudera require it?

Cluster registration is a mandatory process that accompanies Version 2 (V2) license keys. It associates your running clusters with your active Cloudera software license, validates your Cloudera Manager cluster deployments against

your purchased software entitlements and links your active environments directly to your Cloudera Account 360 profile for accurate license tracking.

Which releases introduce mandatory cluster registration?

Cloudera introduces mandatory cluster registration in Cloudera Manager 7.13.2 SP1 (7.13.2.10000) as part of the Cloudera Base on premises 7.3.2 SP1 release line.

When does Cluster Registration become mandatory?

Mandatory registration applies whenever you apply a V2 license key in Cloudera Manager 7.13.2 SP1 (7.13.2.10000) or higher, either during new cluster creation or when updating (or renewing) a license for existing clusters.

What happens to your existing license when you upgrade?

Cloudera Manager converts existing Version 1 (V1) licenses to Version 2 (V2) licenses upon upgrade. Applying a V2 license displays the new **Cluster Registration Status** section in the Web Console and requires you to register all managed clusters.

When does the 60-day grace period start?

The 60-day registration grace period begins on the exact date and time you create a cluster or apply a V2 license key inside the Cloudera Manager Admin Console. It does not begin on the date Cloudera issues or emails the license file.

What happens if Cloudera Manager places a cluster into a *Locked state* due to non-registration?

If you do not register a cluster within the 60-day grace period, or if you apply a V2 license to an existing cluster after its 60-day window ends, Cloudera Manager places the cluster into a *Locked state*. In this state, Cloudera Manager enforces the following restrictions:

- You cannot start or restart any cluster, service, or role.
- You cannot save or deploy configuration changes across the cluster.
- Cloudera Manager locks administrative management functions in the Web Console.

Do running workloads continue processing data if a cluster enters a Locked state?

Yes. Existing running workloads (such as active YARN jobs, Hive queries, or HDFS storage operations) continue to process data normally. However, because Cloudera Manager freezes administrative control, you cannot make configuration adjustments, execute administrative tasks, or restart failed services until you complete registration.

How do I unlock a cluster that has entered the *Locked state*?

Completing the cluster registration process immediately unlocks Cloudera Manager and restores full administrative management functionality.

Do air-gapped clusters require an active internet connection to register?

No. Cloudera Manager supports two workflow options:

- Online Registration: Cloudera Manager automatically validates and registers your clusters in a single click if the system has direct outbound internet access to Cloudera Servers.
- Offline Registration: Cloudera Manager uses a secure, two-way file exchange for air-gapped environments without requiring direct outbound internet connectivity from your clusters:

  1. Export a JSON registration package from your air-gapped Cloudera Manager instance.
  2. Upload the JSON file to Cloudera Account 360 using an internet-connected workstation.
  3. Download the signed confirmation package from Cloudera Account 360 and import it back into Cloudera Manager.

Do I need to register air-gapped clusters individually?

No. Cloudera Manager provides a Bulk Cluster Key Migration feature. You can export a single JSON package containing details for all managed clusters, upload it once to Cloudera Account 360 using an internet-connected device, and import a single confirmation package back to register all clusters simultaneously.

Does the registration JSON package contain sensitive cluster data?

No. The JSON package contains only non-sensitive entitlement identifiers (such as cluster UUIDs, host counts, and hardware specifications) required to verify software license allocations. You can open and inspect the JSON file prior to uploading it to Cloudera Account 360.

What steps do I take when applying a new or renewed license key?

Anytime you apply an updated or renewed V2 license key to Cloudera Manager, you must re-register all active clusters to associate them with the new key:

- Connected Environments: Cloudera Manager automatically contacts Cloudera servers and re-registers all managed clusters.
- Air-Gapped Environments: Administrators must perform a quick offline bulk registration process using Cloudera Account 360.

Why does the Cluster Deregistration page omit active clusters?

The **Cluster Deregistration** page lists only clusters that you have already deleted from Cloudera Manager. Active clusters do not appear on this page until you decommission and remove them.

Why must you deregister a cluster after deleting it?

Deregistering a deleted cluster notifies the Cloudera license server that the cluster is no longer active. This action releases its allocated license entitlements, making those node allocations available for future cluster deployments.

Can I deregister a cluster in an air-gapped environment?

Yes. You can perform offline deregistration by clicking the Offline Deregistration button on the **Cluster Deregistration** page and following the file-transfer workflow with Cloudera Account 360.

How do I remove and deregister a decommissioned cluster from my account?

- Connected Environments: Deleting a cluster in Cloudera Manager automatically sends a deregistration request to Cloudera servers and frees up the resources.
- Air-Gapped Environments: Select the deleted cluster on the **Cluster Deregistration** page, click Offline Deregistration, process the request JSON in Cloudera Account 360, and import the signed confirmation back into Cloudera Manager.

Related Information

Mandatory Cluster Registration Policy

Registering and Deregistering Clusters in Cloudera Manager

Cloudera Manager User Roles

Access to Cloudera Manager features is controlled by user accounts that specify an authentication mechanism and one or more user roles

User roles determine the tasks that an authenticated user can perform and the features visible to the user in the Cloudera Manager Admin Console. In addition to the default user roles, you can create user roles that apply only to specific clusters.

Documentation for Cloudera Manager administration and management tasks indicate user roles required to perform the task.

Important: All possible user roles are available with Cloudera Enterprise. When a Cloudera Enterprise Data Hub Edition trial license expires, only users with Read-Only and Full Administrator roles can log in to Cloudera Manager. A Full Administrator must change user accounts with other roles to Read-Only or Full Administrator before such users can log in.

Important:

Even though the Read-Only role is the least privileged role, it is still allowed to download client configuration files. Client configuration files contain the same information that is visible to the Read-Only role.

Displaying Your Roles

To view your roles, perform the following step:

1. In the Cloudera Manager Admin Console, select <username>My Profile.

Default User Roles

By default, Cloudera Manager ships with user roles that have privileges for all clusters managed by Cloudera Manager. You can create roles that are a combination of a default user role and privileges on a specific cluster. For more information about this type of role, see User Roles with Privileges for a Cluster on page 119.

The following table describes the actions each user role can perform:

Permitted Operations	Auditor	Cluster Administrator	Cluster Creator	Configurator	Dashboard User	Full Administrator	Key Administrator	Limited Cluster Administrator	Limited Operator	Navigator Administrator	Operator	Read-Only	Replication Administrator	User Administrator
Access all functionality that Cloudera Manager offers		Y				Y								
Add and Remove Entity Tags		Y				Y		Y						
Administer Cloudera Navigator		Y				Y				Y				
Apply policies to redact sensitive data		Y				Y								
Configure HDFS Encryption, administer Key Trustee Server, and manage encryption keys						Y	Y							
Create clusters		Y	Y			Y								
Create replication policies and snapshot policies						Y							Y	
Create, modify, and delete your own dashboards					Y	Y								
Create, update, or delete external account configuration						Y								Y
Decommission hosts		Y		Y		Y		Y	Y		Y			
Edit the configuration of services and roles		Y		Y		Y		Y						
Enter and exit Maintenance Mode		Y		Y		Y		Y						
Import Cluster Template		Y				Y		Y						
Inspect Hosts		Y				Y		Y						
Manage Full Administrator accounts						Y								
Manage user accounts and configuration of external authentication						Y								Y
Recommission hosts, and decommission and recommission roles		Y		Y		Y		Y			Y			
See available hosts		Y	Y			Y		Y						
Send Diagnostic Bundles		Y				x		Y						
Start, stop, and restart KMS		Y		Y		Y	Y	Y			Y			
Start, stop, and restart most clusters, services, and roles		Y		Y		Y		Y			Y			
Upgrade Clusters		Y				Y								
View and perform parcels operations		Y	Y			Y		Y						
View audit events	Y					Y				Y				
View data in Cloudera Manager	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Disk Usage By Directory	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Directory Usage		Y				Y								
File Browser		Y				Y								
Kill applications in YARN		Y				Y								
Cancel memory-intensive Impala queries		Y				Y								

User Roles with Privileges for a Cluster

In addition to the default user roles, you can create user roles that apply only to specific clusters. Creating this new role is done by assigning a privilege for a specific cluster to a default role. When a user account has multiple roles, the privileges are the union of all the roles.

For example, the user account milton has the Limited Operator role and Read-Only role with a scope of Cluster 1. Additionally, milton has the Configurator role on Cluster 2.

On Cluster 1, milton can perform all the actions that a Limited Operator and Read-Only can.

On Cluster 2, milton can perform all the actions that a Configurator can.

The user account milton cannot perform these or any other actions on the other clusters that are managed by Cloudera Manager because the account does not have any other roles.

Another user account, edith, has the Configurator role with privileges for all clusters. This means that edith can perform the actions of the Configurator role on all clusters that Cloudera Manager manages since the scope is all clusters.

You can assign privileges for a specific cluster to the following user roles:

- Cluster Administrator
- Configurator
- Limited Operator
- Operator
- Read-Only

User roles that cannot be assigned privileges for a specific cluster apply to all clusters. For example, if edith has the Key Administrator user role, she can perform the actions of a Key Administrator on all clusters.

Figure 1: Cluster-Specific Permissions

Watch the video to set up permissions for a specific cluster at https://youtu.be/9TxpqWaDA500

Adding a User Role for a Specific Cluster

To create a role that has privileges for a specific cluster, perform the following steps:

1. In the Cloudera Manager Admin Console, navigate to AdministrationUsers & RolesRoles.
2. Click Add Role.
3. Specify the following:
   - Privilege: The user role and cluster you want to assign privileges for.
   - Users: The users you want to assign to this new role. You can assign users now or at a later time.
   - LDAP Group/External Program Exit Codes/SAML Attributes/SAML Script Exit Codes: The external mapping you want to assign this new role to. You can assign external mappings now or at a later time with the process described in Mapping External Authentication to a Role on page 121.

     This field is based on your authentication mode and does not appear for local users.

     Valid values for the External Program Exit Code and SAML Script Exit Code are between 0 and 127. You defined what users you want to associate with theses values when you configure your external authentication. For more information,

     If you are upgrading to Cloudera Manager 6 from Cloudera Manager 5, existing mappings are imported from Cloudera Manager 5. These imported mappings can be changed.

     The following list describes the LDAP groups imported from Cloudera Manager 5:

     - LDAP Full Administrator Groups
     - LDAP User Administrator Groups
     - LDAP Cluster Administrator Groups
     - LDAP Replication Administrator Groups
     - LDAP Configurator Groups

- LDAP Key Administrator Groups
- LDAP Navigator Administrator Groups
- LDAP Operator Groups
- LDAP Limited Operator Groups
- LDAP Auditor Groups

The following list describes the SAML and External Program codes imported from Cloudera Manager 5:

- 0 - Full Administrator
- 1 - Read-Only
- 2 - Limited Operator
- 3 - Operator
- 4 - Configurator
- 5 - Cluster Administrator
- 6 - Replication Administrator
- 7 - Navigator Administrator
- 8 - User Administrator
- 9 - Auditor
- 10 - Key Administrator
- 11 - Dashboard User

4. Click Add.

Mapping External Authentication to a Role

If you are using an external authentication, such as a SAML Script, you must map its information to Cloudera Manager user roles. Before you can map a role though, make sure that it exists. If it does not exist, create it by completing the steps described in Adding a User Role for a Specific Cluster on page 120.



Attention: If you do not map an External Authentication entity (such as an LDAP group) to a role, users that belong to that group will default to no access.

For example, you are using a SAML Script and want to assign user accounts that correspond with exit code 15 to a Cluster Administrator role with privileges for a cluster named cluster1.

To accomplish this, perform the following steps in the Cloudera Manager Admin Console:

1. Navigate to AdministrationUsers & RolesRoles.
2. Based on your authentication method, select LDAP Groups, SAML Attributes, SAML Scripts, or External Programs.
3. Click Add <authentication method> Mapping.
4. Fill in the value for your authentication method, such as SAML Script Exit Code, and select the role you want to map to that value from the dropdown menu.

   For SAML Scripts and External Programs, valid values are between 0 and 127.
5. Click Save.
6. Repeat this process for all the roles you want to map.

If you are upgrading to Cloudera Manager 6 from Cloudera Manager 5, existing mappings are imported from Cloudera Manager 5. These imported mappings can be changed.

The following list describes the LDAP groups imported from Cloudera Manager 5:

- LDAP Full Administrator Groups
- LDAP User Administrator Groups
- LDAP Cluster Administrator Groups
- LDAP Replication Administrator Groups
- LDAP Configurator Groups

- LDAP Key Administrator Groups
- LDAP Navigator Administrator Groups
- LDAP Operator Groups
- LDAP Limited Operator Groups
- LDAP Auditor Groups

The following list describes the SAML and External Program codes imported from Cloudera Manager 5:

- 0 - Full Administrator
- 1 - Read-Only
- 2 - Limited Operator
- 3 - Operator
- 4 - Configurator
- 5 - Cluster Administrator
- 6 - Replication Administrator
- 7 - Navigator Administrator
- 8 - User Administrator
- 9 - Auditor
- 10 - Key Administrator
- 11 - Dashboard User

Assigning Users to a Role

In addition to mapping groups, such as LDAP groups, to a user role, you can also assign individual users to a user role. If you do not assign a role, the local user defaults to no access. This means that the user cannot perform any actions on the cluster.

To add a user account to a role, perform the following steps:

1. In the Cloudera Manager Admin Console, navigate to AdministrationUsers & RolesRoles.
2. Click Assign for the role you want to modify.
3. Specify the Users or <Authentication Method Value> groups you want to assign to the role.
4. Save the changes.

Removing a User or External Mapping from a User Role

Perform the following steps to remove a user account or external mapping from a user role:

1. In the Cloudera Manager Admin Console, navigate to AdministrationUsers & RolesRoles.
2. Click Assign for the role you want to modify.
3. Click the X for each user or external mapping you want to remove from the user role and click Save.

Removing a Role

To remove a role with a specific privilege, you must first remove all the user accounts that have that role. Note that you cannot remove the default roles that Cloudera Manager ships with.

The following steps describe how to remove users and then delete the role:

1. In the Cloudera Manager Admin Console, navigate to AdministrationUsers & RolesRoles.
2. Click Assign for the role you want to modify.
3. Click the X for each user or external mapping you want to remove from the user role and click Save.
4. Click Remove.

Removing the Full Administrator User Role

Minimum Required Role: User Administrator (also provided by Full Administrator) This feature is not available when using to manage Data Hub clusters.

In some organizations, security policies may prohibit the use of the Full Administrator role. The Full Administrator role is created during Cloudera Manager installation, but you can remove it as long as you have at least one remaining user account with User Administrator privileges.

To remove the Full Administrator user role, perform the following steps.

1. Add at least one user account with User Administrator privileges, or ensure that at least one such user account already exists.
2. Ensure that there is only a single user account with Full Administrator privileges.
3. While logged in as the single remaining Full Administrator user, select your own user account and either delete it or assign it a new user role.



Warning: After you delete the last Full Administrator account, you will be logged out immediately and will not be able to log in unless you have access to another user account. Also, it will no longer be possible to create or assign Full Administrators.

A consequence of removing the Full Administrator role is that some tasks may require collaboration between two or more users with different user roles. For example:

- If the machine that the Cloudera Navigator roles are running on needs to be replaced, the Cluster Administrator will want to move all the roles running on that machine to a different machine. The Cluster Administrator can move any non-Navigator roles by deleting and re-adding them, but would need a Navigator Administrator to perform the stop, delete, add, and start actions for the Cloudera Navigator roles.
- In order to take HDFS snapshots, snapshots must be enabled on the cluster by a Cluster Administrator, but the snapshots themselves must be taken by a Replication Administrator.

Other Cloudera Manager Tasks and Settings

From the Administration tab you can select options for configuring settings that affect how Cloudera Manager interacts with your clusters.

Settings

The Settings page provides a number of categories as follows:

- Performance - Set the Cloudera Manager Agent heartbeat interval.
- Advanced - Enable API debugging and other advanced options.
- Monitoring - Set Agent health status parameters.
- Security - Set TLS encryption settings to enable TLS encryption between the Cloudera Manager Server, Agents, and clients. You can also:
  - Set the realm for Kerberos security and point to a custom keytab retrieval script.
  - Specify session timeout and a "Remember Me" option.
- Ports and Addresses - Set ports for the Cloudera Manager Admin Console and Server.
- Other
  - Enable Cloudera usage data collection.
  - Set a custom header color and banner text for the Admin console.
  - Set an "Information Assurance Policy" statement – this statement will be presented to every user before they are allowed to access the login dialog box. The user must click "I Agree" in order to proceed to the login dialog box.
  - Disable/enable the auto-search for the Events panel at the bottom of a page.
- Support
  - Configure diagnostic data collection properties.
  - Configure how to access Cloudera Manager help files.

- External Authentication - Specify the configuration to use LDAP, Active Directory, or an external program for authentication.
- Parcels - Configure settings for parcels, including the location of remote repositories that should be made available for download, and other settings such as the frequency with which Cloudera Manager will check for new parcels, limits on the number of downloads or concurrent distribution uploads. See Parcels for more information.
- Network - Configure proxy server settings.
- Custom Service Descriptors - Configure custom service descriptor properties for Add-on services.

Alerts

See *Managing Alerts*.

Users

See *Cloudera Manager User Accounts*.

Kerberos

See *Enabling Kerberos Authentication for Cloudera Runtime*.

License

See *Managing Licenses*.

User Interface Language

You can change the language of the Cloudera Manager Admin Console User Interface through the language preference in your browser. Information on how to do this for the browsers supported by Cloudera Manager is shown under the Administration page. You can also change the language for the information provided with activity and health events, and for alert email messages by selecting Language, selecting the language you want from the drop-down list on this page, then clicking Save Changes.

Peers

See *Designating a Replication Source*.

Providing full administrative permissions to the users on a temporary basis

When users login to Cloudera Manager UI, there are only limited admins, which implies that they are restricted to perform several modifications like adding roles to hostgroups and pertinent operations. On a short term basis, it could be necessary for the users to fix a cluster or perform some immediate task(s) that can unblock performance related challenges.

About this task

The users are provided access to perform short term operational tasks and improve the overall functioning of Cloudera Manager.

Procedure

1. SSH to the Cloudera Manager instance and switch to root user.
2. Grab the DB access info from:

```
/etc/cloudera-scm-server/db.properties

    1. You will need the following values:

com.cloudera.cmf.db.host
```

```
com.cloudera.cmf.db.name

com.cloudera.cmf.db.user

com.cloudera.cmf.db.password

    2. You can use the following command to export them to variables. Use
 PGPASSWORD to avoid password prompt:

export DBHOST=$(grep "db.host" /etc/cloudera-scm-server/db.properties |
cut -d'=' -f2);

export DBNAME=$(grep "db.name" /etc/cloudera-scm-server/db.properties | cu
t -d'=' -f2);

export DBUSER=$(grep "db.user" /etc/cloudera-scm-server/db.properties |
cut -d'=' -f2);

export PGPASSWORD=$(grep "db.password" /etc/cloudera-scm-server/db.propert
ies | cut -d'=' -f2);
```

3. Connect to the database:

   psql -d $DBNAME -h $DBHOST -U $DBUSER

4. Adding user as admin:

   a. Get user ID (user must have logged on previously):

      select USER_ID,USER_NAME from USERS;

   b. Set admin role (replace <user's id> with the right one from the previous select):

      INSERT INTO user_auth_roles SELECT <user's id>, auth_role_id FROM auth_roles WHERE a
      uth_roles.name='ROLE_ADMIN';

   c. Quit postgres

      Use CTRL+d or \q

5. Restart Cloudera Manager.

   For the changes to take effect immediately, you might have to restart Cloudera Manager:
   systemctl restart cloudera-scm-server.service

What to do next

Be sure you have cleaned up cookies in your browser or use a new private window when trying to login using the new role.

Cloudera Management Service

The Cloudera Management Service is a set of roles used by Cloudera Manager to manage and monitor clusters.

The Cloudera Management Service implements various management features as a set of roles:

- Host Monitor - collects health and metric information about hosts
- Service Monitor - collects health and metric information about services and activity information from the YARN and Impala services
- Event Server - aggregates relevant Hadoop events and makes them available for alerting and searching
- Alert Publisher - generates and delivers alerts for certain types of events
- Telemetry Publisher - collects and sends workload information to Cloudera Observability. For example, when new clusters are added with Cloudera Manager, Telemetry Publisher automatically sends the new cluster information to Cloudera Observability.

- Reports Manager - generates reports that provide an historical view into disk utilization by user, user group, and directory, processing activities by user and YARN pool, and HBase tables and namespaces. This role is not added in Cloudera Express.

You can view the status of the Cloudera Management Service by doing one of the following:

- Select Clusters Cloudera Management Service .
- On the Home Status tab, in Cloudera Management Service table, click the Cloudera Management Service link.

Health Tests

Cloudera Manager monitors the health of the services, roles, and hosts that are running in your clusters using *health tests*. The Cloudera Management Service also provides health tests for its roles. Role-based health tests are enabled by default. For example, a simple health test is whether there's enough disk space in every NameNode data directory. A more complicated health test may evaluate when the last checkpoint for HDFS was compared to a threshold or whether a DataNode is connected to a NameNode. Some of these health tests also aggregate other health tests: in a distributed system like HDFS, it's normal to have a few DataNodes down (assuming you've got dozens of hosts), so we allow for setting thresholds on what percentage of hosts should color the entire service down.

Health tests can return one of three values: Good, Concerning, and Bad. A test returns Concerning health if the test falls below a warning threshold. A test returns Bad if the test falls below a critical threshold. The overall health of a service or role instance is a roll-up of its health tests. If any health test is Concerning (but none are Bad) the role's or service's health is Concerning; if any health test is Bad, the service's or role's health is Bad.

In the Cloudera Manager Admin Console, health tests results are indicated with colors: Good , Concerning , and Bad .

One common question is whether monitoring can be separated from configuration. One of the goals for monitoring is to enable it without needing to do additional configuration and installing additional tools (for example, Nagios). By having a deep model of the configuration, Cloudera Manager is able to know which directories to monitor, which ports to use, and what credentials to use for those ports. This tight coupling means that, when you install Cloudera Manager all the monitoring is enabled.

Metric Collection and Display

To perform monitoring, the Service Monitor and Host Monitor collects metrics. A *metric* is a numeric value, associated with a name (for example, "CPU seconds"), an entity it applies to ("host17"), and a timestamp. Most metric collection is performed by the Agent. The Agent communicates with a supervised process, requests the metrics, and forwards them to the Service Monitor. In most cases, this is done once per minute.

A few special metrics are collected by the Service Monitor. For example, the Service Monitor hosts an HDFS canary, which tries to write, read, and delete a file from HDFS at regular intervals, and measure whether it succeeded, and how long it took. Once metrics are received, they're aggregated and stored.

Using the Charts page in the Cloudera Manager Admin Console, you can query and explore the metrics being collected. Charts display *time series*, which are streams of metric data points for a specific entity. Each metric data point contains a timestamp and the value of that metric at that timestamp.

Some metrics (for example, total_cpu_seconds) are counters, and the appropriate way to query them is to take their rate over time, which is why a lot of metrics queries contain the dt0 function. For example, dt0(total_cpu_seconds). (The dt0 syntax is intended to remind you of derivatives. The `0` indicates that the rate of a monotonically increasing counter should never have negative rates.)

Events, Alerts, and Triggers

An *event* is a record that something of interest has occurred – a service's health has changed state, a log message (of the appropriate severity) has been logged, and so on. Many events are enabled and configured by default.

An *alert* is an event that is considered especially noteworthy and is triggered by a selected event. Alerts are shown with an  badge when they appear in a list of events. You can configure the Alert Publisher to send alert notifications by email or by SNMP trap to a trap receiver.

A *trigger* is a statement that specifies an action to be taken when one or more specified conditions are met for a service, role, role configuration group, or host. The conditions are expressed as a tsquery statement, and the action to be taken is to change the health for the service, role, role configuration group, or host to either Concerning (yellow) or Bad (red).

Starting the Cloudera Management Service

How to start the Cloudera Management Service.

Before you begin

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

Procedure

1. Do one of the following:
   - **a.** Select Clusters Cloudera Management Service .
     b. Select ActionsStart.
   - On the HomeStatus tab, click the options menu to the right of Cloudera Management Service and select Start.
2. Click Start to confirm. The Command Details window shows the progress of starting the roles.

Results

When Command completed with *N*/*N* successful subcommands appears, the task is complete. Click Close.

Stopping the Cloudera Management Service

How to stop the Cloudera Management Service.

Before you begin

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

Procedure

1. Do one of the following:
   - **a.** Select Clusters Cloudera Management Service .
     b. Select ActionsStop.
   - On the HomeStatus tab, click the options menu to the right of Cloudera Management Service and select Stop.
2. Click Stop to confirm. The Command Details window shows the progress of stopping the roles.

Results

When Command completed with *N*/*N* successful subcommands appears, the task is complete. Click Close.

Restarting the Cloudera Management Service

How to restart the Cloudera Management Service.

Before you begin

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

Procedure

1. Do one of the following:
   - **a.** Select Clusters Cloudera Management Service .
     b. Select ActionsRestart.
   - On the HomeStatus tab, click the options menu to the right of Cloudera Management Service and select Restart.
2. Click Restart to confirm. The Command Details window shows the progress of restarting the roles.

Results

When Command completed with *N*/*N* successful subcommands appears, the task is complete. Click Close.

Starting and Stopping Cloudera Management Service Roles

Before you begin

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

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Procedure

1. Do one of the following:
   - Select Clusters Cloudera Management Service .
   - On the HomeStatus tab, in Cloudera Management Service table, click the Cloudera Management Service link.
2. Click the Instances tab.
3. Select a role.
4. Do one of the following:s
   - Start: Select Actions for SelectedStart and click Start to confirm
   - Stop: Select Actions for SelectedStop and click Stop to confirm.

Results

When Command completed with *N*/*N* successful subcommands appears, the task is complete. Click Close.

Configuring Management Service Database Limits

Configuring database service limits lets you control the amount of retained monitoring data.

Before you begin

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

About this task

Each Cloudera Management Service role maintains a database for retaining the data it monitors. These databases (as well as the log files maintained by these services) can grow quite large. Limits on these data sets are configured when you create the management services, but you can modify these parameters through the Configuration settings in the Cloudera Manager Admin Console. For example, the Event Server lets you set a total number of events to store.

There are also settings for the logs that these various services create. You can throttle how big the logs are allowed to get and how many previous logs to retain.

Procedure

1. Do one of the following:
   - Select Clusters Cloudera Management Service .
   - On the HomeStatus tab, in Cloudera Management Service table, click the Cloudera Management Service link.
2. Click the Configuration tab.
3. Select Scope and then one of the following.
   - Host Monitor
   - Service Monitor
4. Select CategoryLog Files to view log file size properties.
5. Edit the appropriate properties.

   To apply this configuration property to other role groups as needed, edit the value for the appropriate role group. See .
6. Click Save Changes.

Related Information

Data Storage for Monitoring Data

Configuring Cloudera Manager for High Availability

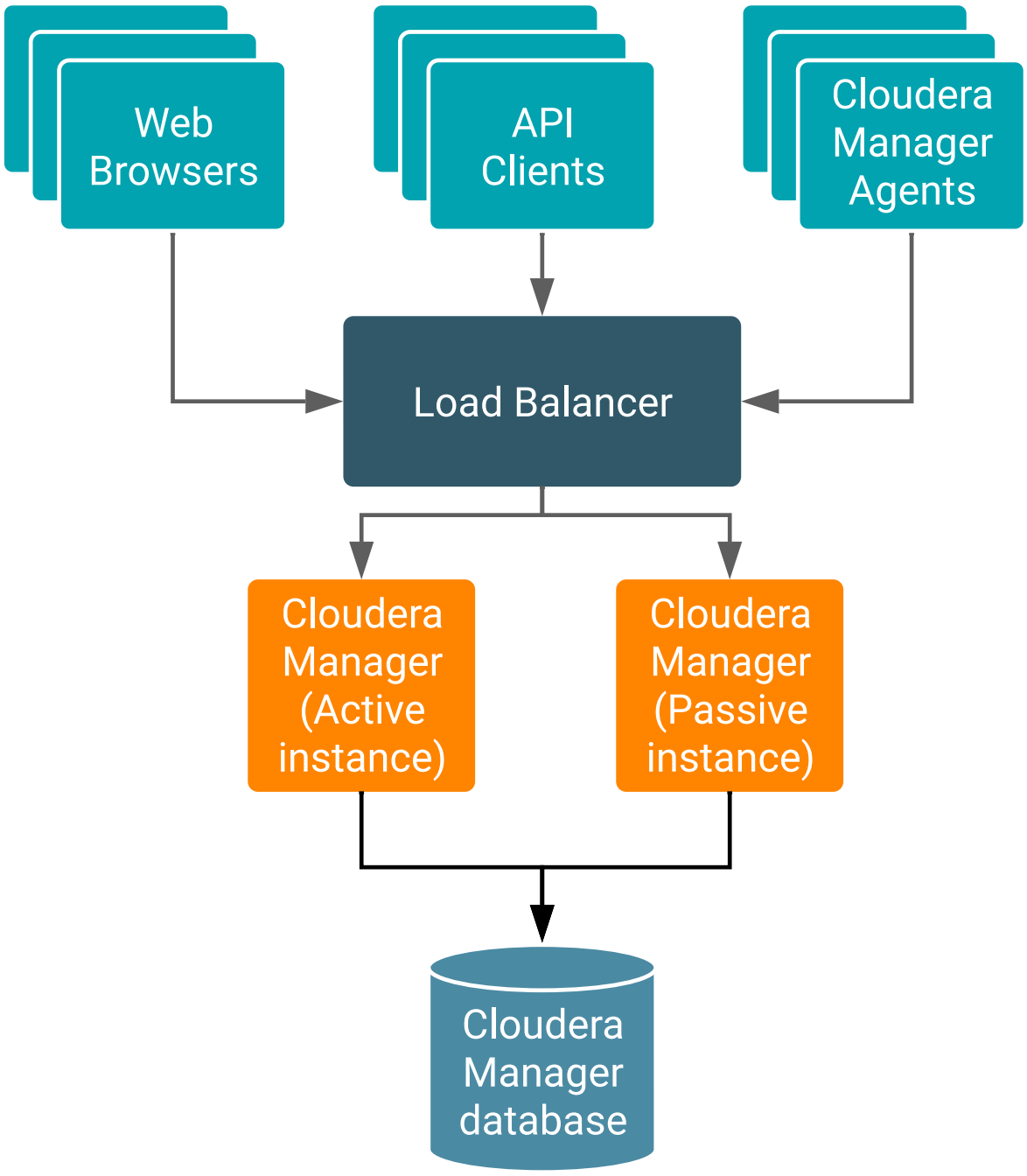
You can configure Cloudera Manager for high availability. This configuration provides failover capability in case an instance of Cloudera Manager fails.



Note: Cloudera Manager high availability feature is only supported with CDP Private Cloud Base 7.1.8 version or higher.

These steps configure Cloudera Manager into an highly available Active-Passive configuration. The Active-Passive configuration supports only 2 hosts running Cloudera Manager along with an external load balancer. Load balancer configuration defines the active vs passive hosts and active host serves all requests during normal operation. The load balancer monitors both active and passive hosts and automatically diverts requests to the passive host if the active host is unresponsive for a given (configured in the load balancer) period of time. If the active host becomes available again, the load balancer automatically redirects the requests back to the active host. Cloudera Manager also internally uses a priority parameter to decide which host will take the active role to process commands within the system and which host will take over the passive role when both hosts boot up and become functional.

Figure 2: Architecture



Prerequisites for configuring Cloudera Manager for high availability

Prerequisites for configuring Cloudera Manager for high availability.

Important: If you are on an old Cloudera Manager high availability setup and planning to upgrade to Cloudera Manager 7.7.1 or higher version then post Cloudera Manager upgrade, you must remove the old Cloudera Manager high availability setup and set up a new high availability for Cloudera Manager 7.7.1 or higher version as per the following instructions. The old Cloudera Manager high availability setup is not supported with Cloudera Manager 7.7.1 or higher version.

To configure high availability for Cloudera Manager, you must:

- Provide an separate host for each Cloudera Manager instance.
- Configure the database used by Cloudera Manager to run on a different host than the Cloudera Manager instances.
- Install and configure a load balancer to process requests to Cloudera Manager.
- If you are expanding an existing deployment to use Cloudera Manager high availability:
  - Backup the Cloudera Manager database. See Back Up the Cloudera Manager Databases./
  - If TLS is enabled, backup the certificates.
- The load balancer must be configured to accept connection requests (both TCP and HTTP) for Cloudera Manager and redirect them to one of the Cloudera Manager instances. Consult the documentation for your load balancer for setup and configuration procedures.

  Cloudera recommends the HAProxy load balancer. See Sample Configuration Files for HAProxy on page 137.
- If TLS is enabled for Cloudera Manager, configure the load balancer as follows:
  - The load balancer must be configured with its own TLS certificate.
  - The load balancer must be configured to terminate and re-establish TLS connections to the underlying Cloudera Manager hosts on ports 7182 and 7183. TLS must be enabled on those ports.
  - The load balancer configuration for port 7182 must use the load balancer's certificate for mutual TLS authentication.
  - If using Auto-TLS, all Cloudera Manager Server hosts must also have the Cloudera Manager agent package installed on them.
- If using Kerberos Active Directory for user authentication, ensure that the Kerberos and LDAP client packages are installed on all Cloudera Manager hosts.

User impacts and limitations for Cloudera Manager in high availability mode

There are impacts and limitations you should be aware of when using Cloudera Manager in high-availablity mode.

Managing Cloudera Manager Wizards in High Availability (HA) mode

Problem Statement

In a Cloudera Manager High Availability (HA) environment, both Cloudera Manager Server instances monitor the same shared parcel repository and share a single PostgreSQL database.

When adding a new parcel URL or uploading a parcel, both nodes might attempt to process the metadata simultaneously, causing the parcel addition to fail and triggering a following unique constraint violation in the Cloudera Manager Server log:

```
Detail: Key (product, version)=(CDH, 7.1.9...) already exists.
```



Caution: Do not trigger the Cloudera Manager Wizard for activities such as installation, upgrades, adding Runtime services, or Cloudera Data Services on premises deployment while you have Cloudera Manager configured for HA.

Root Cause

The issue is caused by a race condition. Because both Cloudera Manager instances are active and reading from and writing to the same database schema, they might simultaneously attempt to insert the same parcel record. This leads to database locks or "duplicate key" errors.

This same principle applies to Cloudera Manager Wizard activities (such as installation, upgrades, adding Runtime services, or Cloudera Data Services on premises deployment), where concurrent DB writes can lead to inconsistent state or task failures.

Recommended Solution: Single-Node Maintenance Mode

To ensure database integrity during parcel operations or wizard executions, you must temporarily transition to a single-node configuration.

1. Prepare the Primary Instance
   a. Stop the secondary Cloudera Manager Server instance.
   b. On the primary Cloudera Manager node, open the /etc/default/cloudera-scm-server configuration file.
   c. Locate the CMF_JAVA_OPTS variable and modify the haMode flag to false:

      ```
      -Dcom.cloudera.cmf.haMode=false
      ```

   d. Restart the Cloudera Manager Server on the primary node only.
2. Perform Maintenance
   a. Log in to the Cloudera Manager UI.
   b. Add the Parcel URL, download or distribute parcels, or run the desired Wizard (Install, Upgrade, or Add Runtime service).
   c. Verify that the operation has completed successfully.
3. Restore High Availability
   a. On the primary Cloudera Manager node, revert the haMode flag in /etc/default/cloudera-scm-server to true:

      ```
      -Dcom.cloudera.cmf.haMode=true
      ```

   b. Restart the Cloudera Manager Server on the primary node.
   c. Start the Cloudera Manager Server on the secondary node.
   d. Confirm both nodes are synchronized, and the HA status is healthy in the Cloudera Manager UI.

Important Restrictions

While in HA mode, do not perform the following activities unless you follow the "Single-Node" steps described above:

- Adding/Updating Parcels: This prevents contention for locks on the RELEASES table.
- Cluster Upgrades: This ensures that a single orchestrator manages the upgrade path.
- Adding Runtime services: This prevents the system from generating duplicate service IDs or creating configuration conflicts.
- Cloudera Data Services on premises Deployment: This ensures that environment validation runs on a single thread.



Important:

This behavior affects all Cloudera Runtime versions when you configure Cloudera Manager for HA with a shared PostgreSQL 16 (or similar) backend.

Known Limitations

For a cluster running with Cloudera Manager High Availability in Active-Passive, it has some known limitations that you must be aware of.

- During the failover time (normally, a few seconds), any incoming API calls will be dropped because the load balancer will not redirect any API call to the Passive server during that period.
- Some commands that work with files (e.g. GenerateCMCA) may fail if a Cloudera Manager Server failover happens halfway. Those files will not be transferred during failover. You should retry such commands manually.
- You cannot enable Secure Credential Storage if Cloudera Manager high availability has been configured.
- The QueueManager service should be manually restarted if it indicates stale configuration after you restart the Cloudera Manger server after setting up High Availability.

- If the Active instance of Cloudera Manager server fails, and Cloudera Manager fails over to the Passive instance, Parcels must be downloaded again to the Passive instance. (Required only when adding new hosts, or services.)
- You can see some instances of following errors in server logs, which can be safely ignored:

```
ERROR org.hibernate.engine.j1dbc.batch.internal.BatchingBatch - HHH00031
5:
Exception executing batch [org.hibernate.StaleStateException:
Batch update returned unexpected row count from update [0]; actual row co
unt: 0; expected: 1; statement executed:
```

- Hive Replication may fail if it occurs during a Cloudera Manager failover event. After the failover, subsequent replications will succeed.
- **OPSAPS-60726: Newly saved parcel URLs are not showing up in the parcels page in the Cloudera Manager HA cluster.**

  To safely manage parcels in a Cloudera Manager HA environment, follow these steps:

  1. Shutdown the Passive Cloudera Manager Server.
  2. Add and manage the parcel as usual, as described in Install Parcels.
  3. Restart the Passive Cloudera Manager server after parcel operations are complete.

Cloudera Manager behavior

For a cluster running with Cloudera Manager High Availability in Active-Passive, it may behave differently from a non-HA Cluster in terms of success/failure/retry of operations.

If the Active Cloudera Manager server fails while running commands or API calls:

In general, “a server communication error warning” will pop up on the command processing UI. Once the Load Balancer redirects the traffic to the passive server, the warning will go away and the command will continue running.

The failover time should be fairly short (the default for HAProxy is 4-6 seconds). However, during that period, Cloudera Manager will fail to respond to any API call and does not retry failed API calls.

if the Active server is running while the Passive server fails

If the Passive server node fails, the load balancer should be aware of this by monitoring the health check packets. As long as the Active node is running, users will not be affected.

If both Active and Passive servers fail

The Cloudera Manager Admin Console will no longer be accessible since there is no functional server running at this point.

Once a Cloudera Manager Server node has been repaired or replaced (either the Active or Passive instance), the Load Balancer will start directing traffic to that node, and commands that were in flight at the time of failure will be restarted.

Time taken by failover from the Active to Passive server instance

It depends on the health check settings of your load balancer. Take HAProxy as an example, by default, three consecutive failed checks are needed to remove the server from the load balancing rotation. The interval of each health check is normally 2 seconds. So it might take HaProxy 4-6 seconds to do the failover under a default setting.

Time taken to fail back from the Passive server to the Active server

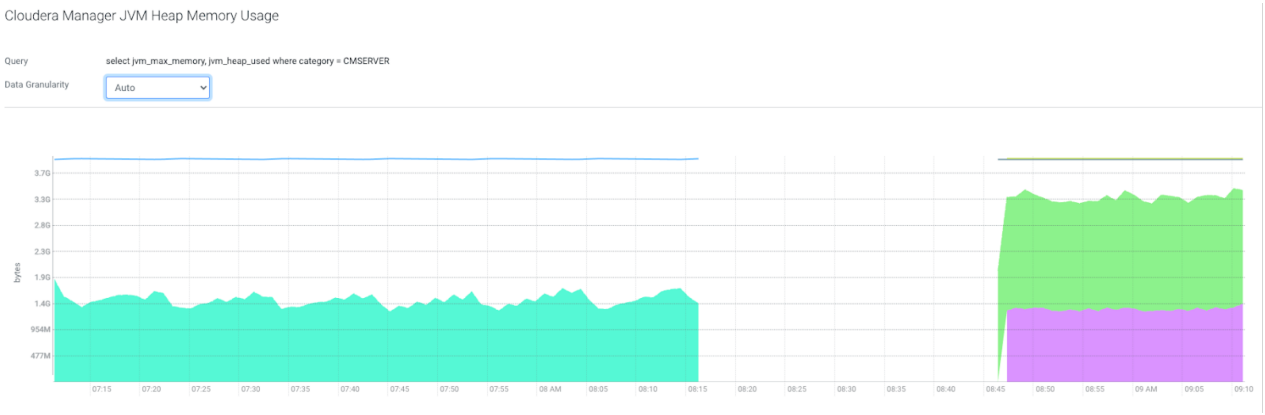
It can take up to 25 seconds to fail back from the Passive server to the Active server. During this time the Cloudera Manager Admin Console returns a 404 error. This additional time is taken by the Cloudera Manager server to become operational.

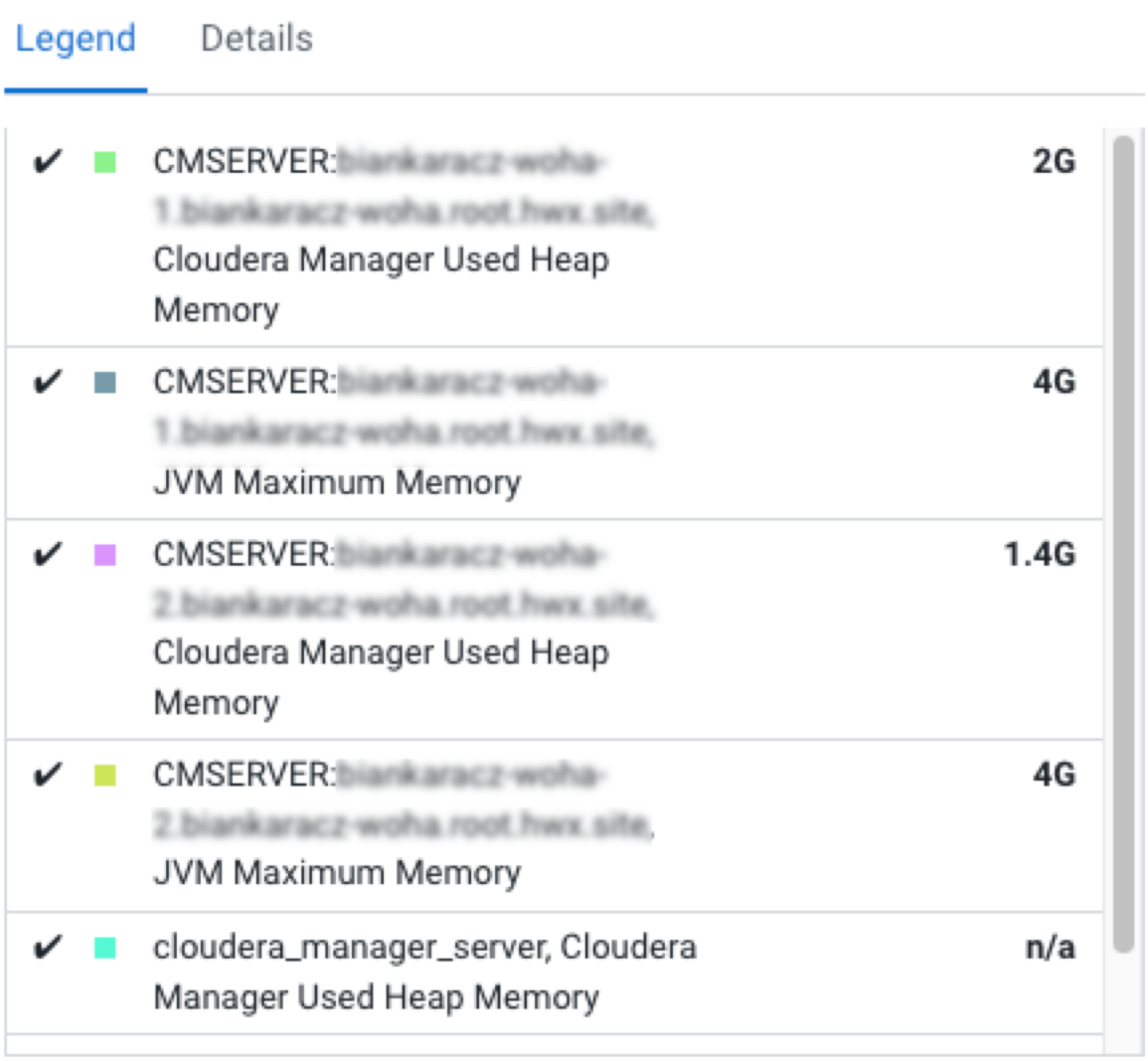
Changes in Health Tests

On HA-mode compatible versions, the naming of the Cloudera Manager server entities has changed to distinguish between multiple instances. This change will impact the customers that are upgrading from an earlier version.

In the earlier versions, the Cloudera Manager server name was cloudera_manager_server, which changed to CMSE RVER:*{CLOUDERA MANAGER SERVER'S HOST NAME}*.Due to this name convention change, you will observe in the chart below that the Cloudera Manager server has a new name and color on the charts. Meanwhile, the metrics for the old entity name are still available.

The cyan area in the chart represents the single Cloudera Manager server running in non-HA mode. The green and purple areas represent the 2 Cloudera Manager servers with HA enabled.





The metrics are still available with the old entity name (cloudera_manager_server) after an upgrade to an HA-compatible version.

The filter for all Cloudera Manager server metrics has been changed. In the earlier versions this query has been used to query cm_database_size:

```
SELECT cm_database_size WHERE entityName = "cloudera_manager_server" AND cat
egory = CMSERVER
```

Now in the predefined plots and chart library, this query is being used:

```
select cm_database_size where category = CMSERVER
```

Installing Cloudera Manager with High Availability

Steps to install Cloudera Manager with high availability.

1. Provision an external database for use by Cloudera Manager.
2. Provision two hosts for Cloudera Manager.
3. Repeat the steps to install Cloudera Manager on each of the hosts to be used for Cloudera Manager high availability. Skip these auto-TLS steps on the passive host. Do not start Cloudera Manager on any hosts until you complete the following:

a. Configure both the active and passive Cloudera Manager instances to connect to the same database. You configure this by running the scm_prepare_database.sh script on each Cloudera Manager host. Specify the database host with the --host option.

b. 1. On each Cloudera Manager host, edit the following file: /etc/default/cloudera-scm-server
   2. Find the line that begins with export CMF_JAVA_OPTS and add the following parameter:

```
-Dcom.cloudera.cmf.haMode=true
```

   3. Find the line that begins with export CMF_SERVER_ARGS and insert the following inside the quotes: --ha-priority <number>

   The --ha-priority sets the priority of the server. The number has to be a non zero positive integer. The lower the number is, the higher priority the server has. Thus, the active server should have a smaller “ha-priority” number than the passive server. When the hardware of the previous running host gets replaced, make sure the new active server always has a higher priority (smaller ha-priority number) than the passive server. To start with, the suggested values are ‘10’ for the active server, and ‘20’ for the passive server. Increment by 10 for each additional passive server.



Important: If you use an incorrect "ha-priority" value, it might cause the Cloudera Manager Load Balancer to use the passive Cloudera Manager server instead of the active Cloudera Manager server for some internal tasks and can lead to Cloudera Manager UI errors such as Command result data not available.

   For example:

```
export CMF_SERVER_ARGS=”--ha-priority 10”
export CMF_JAVA_OPTS="-Xmx2G -XX:MaxPermSize=256m -XX:+HeapDumpOnOutOf
MemoryError  -XX:HeapDumpPath=/tmp -Dcom.cloudera.cmf.haMode=true
```

c. Save the file

4. Start one instance of Cloudera Manager and wait for the Cloudera Manager server to come online.
5. Start the Load Balancer.
6. Configure the Cloudera Manager Hostname Override:

   a. Log in to the Cloudera Manager Admin Console, and go to Administration > Settings.
   b. Select Ports and Addresses.
   c. Set the Cloudera Manager Hostname Override parameter to contain the fully qualified domain name of the load balancer.

7. If you have enabled TLS for Cloudera Manager, do the following:

   a. Locate the Verify Agent Hostname Against Certificate configuration parameter, and uncheck it. (Go to Administration > Settings and search for this parameter.)
   b. Ensure that the CA certificate signing the load balancer’s TLS certificate is trusted by Cloudera Manager server.
   c. The CA certificate signing the load balancer’s TLS certificate must be specified in the file referenced by the Cloudera Manager TLS/SSL Trust Store File configuration parameter (Go to AdministrationSettings Security and search for this parameter.)

8. Restart the active instance of Cloudera Manager Server.
9. CM agents on all hosts must point to load balancer for communicating with CM servers.

   If you are adding high availability to an existing set of hosts managed by Cloudera Manager, configure the Cloudera Manager agents on all managed hosts with the fully-qualified domain name of the load balancer.



Note: If you are setting up a new cluster with new hosts, skip this step. On all existing Cloudera Manager-managed hosts, change the following in the /etc/cloudera-scm-agent/config.ini file:

```
server_host=FULLY QUALIFIED DOMAIN NAME OF THE LOAD BALANCER
```

If your host provisioning system automatically adds new hosts to Cloudera Manager without using the Cloudera Manager Add Host wizard, ensure the above server_host line in the /etc/cloudera-scm-agent/config.ini is configured with the fully-qualified domain name of the load balancer. This may be needed in systems using pre-created host images.

10. If you have enabled TLS for Cloudera Manager, the CA certificate signing the load balancer's TLS certificate must be specified. On all Cloudera Manager-managed hosts, add the following to the /etc/cloudera-scm-agent/config.ini file:

```
verify_cert_file=LOAD_BALANCER_CA_CERT.PEM
```

(Replace load_balancer_ca_cert.pem with the file containing the certificate.)

11. Perform rotation of Auto-TLS certificates (See Rotate Auto-TLS Certificate Authority and Host Certificates) with location parameter set to blank to store the certificates in DB instead of file system. This is important to ensure that both the hosts are able to use the auto TLS certificates after rotation.
12. Restart the Cloudera Manager agents by running the following command on all managed hosts:

```
sudo service cloudera-scm-agent restart
```

13. Start the passive instance of the Cloudera Manager server.
14. Continue with cluster installation and wait for the new cluster to start up.
15. Verify that both Cloudera Manager hosts are showing good health in the Cloudera Manager Admin Console. (Go to Hosts All Hosts to check the status of the hosts of the Cloudera Manager Servers.)

Replacing Cloudera Manager Hosts

Primary/secondary node fails:

1. Set up a new Cloudera Manager server node as per above steps
2. Update the load balancer configuration to update the host name of replaced host.
3. Ensure that the priority of active host remains higher (smaller value of the --ha-priority property) than the passive host.
4. Ensure the Cloudera Manager server is running on the host before restarting load balancer.
5. Restart load balancer server.

Load Balancer fails

1. Set up a new load balancer using the existing configuration.
2. Ensure that the load balancer endpoint remains the same as it is configured in the Cloudera Manager server and agents.

Sample Configuration Files for HAProxy

You can use the supplied configuration files to configure the HAProxy load balancer for deployments with and without TLS or as a guide when using a different type of load balancer.

HAProxy is an open source load balancer. It can be used as a load balancer in front of the Cloudera Manager hosts. Other load balancers may be used instead of HAProxy.

Below is are samples of an haproxy.config file, one for deployments where TLS is enabled, and a second samlple without TLS. If you choose a different type of load balancer, use this configuration file as a guide for configuring the load balancer.

In the files below, <cert.pem> refers to a PEM-encoded file containing a concatenation of the load balancer's certificate and corresponding private key. Ensure that the private key does not have a passphrase. HAProxy does not support private keys protected with a passphrase.

Replace <cm_host_1> and <cm_host_2> with the actual DNS names of the Cloudera Manager server hosts. The passive server is indicated by the use of the backup keyword in the server definition line, as used in the file below.

Sample haproxy.confg file with TLS enabled

```
#---------------------------------------------------------------------
# common defaults that all the 'listen' and 'backend' sections will
# use if not designated in their block
#---------------------------------------------------------------------
defaults
timeout http-request    10s
    timeout queue           1m
    timeout connect         10s
     timeout client          10m
      timeout server          10m
     #timeout http-keep-alive 10s
     timeout check           10s
      maxconn                 3000

# enable admin stats at :8000/haproxy?stats
listen admin
 bind *:8000
 stats enable



global
    log /dev/log local0
    log localhost local1 notice
    maxconn 2000
    daemon

defaults
    log global
    mode tcp
    option tcplog
    option dontlognull
    retries 3
    timeout connect 5000
    timeout client 50000
    timeout server 50000

frontend http-in
    bind *:7183 ssl crt <cert.pem>
    default_backend cmservers

backend cmservers
    mode tcp
    option tcplog
    option log-health-checks
    option redispatch
    log global
    balance roundrobin
    timeout connect 10s
    timeout server 1m
    # active server
    server cm1 <CM_HOST_1>:7183 check ssl verify none crt <cert.pem>
    # passive server
    server cm2 <CM_HOST_2>:7183 check backup ssl verify none crt <cert.pem>


frontend agents
    bind *:7182 ssl crt <cert.pem>
    default_backend foragents

backend foragents
```

```
    mode tcp
    option tcplog
    option log-health-checks
    option redispatch
    log global
    balance roundrobin
    timeout connect 10s
    timeout server 1m
    server cm1 <CM_HOST_1>:7182 check ssl verify none crt <cert.pem>
    server cm2 <CM_HOST_2>:7182 check backup ssl verify none crt <cert.pem>
```

Sample haproxy.confg file for HAProxy without TLS

```
defaults
timeout http-request    10s
     timeout queue           1m
     timeout connect         10s
     timeout client          10m
      timeout server          10m
     #timeout http-keep-alive 10s
     timeout check           10s
      maxconn                 3000

# enable admin stats at :8000/haproxy?stats
listen admin
 bind *:8000
 stats enable


global
    log /dev/log local0
    log localhost local1 notice
    maxconn 2000
    daemon
defaults
    log global
    mode tcp
    option tcplog
    option dontlognull
    retries 3
    timeout connect 5000
    timeout client 50000
    timeout server 50000
frontend http-in
    bind *:7180
    default_backend cmservers
backend cmservers
    mode tcp
    option tcplog
    option log-health-checks
    option redispatch
    log global
    balance roundrobin
    timeout connect 10s
    timeout server 1m
    server cm1 <cm_host_1>:7180 check
    server cm2 <cm_host_2>:7180 check backup
frontend agents
    bind *:7182
    default_backend foragents

backend foragents
```

```
    mode tcp
    option tcplog
    option log-health-checks
    option redispatch
    log global
    balance roundrobin
    timeout connect 10s
    timeout server 1m
    server cm1 <cm_host_1>:7182 check
    server cm2 <cm_host_2>:7182 check backup
```

Securing sensitive information using a Secure Credential Storage Provider

You can set up Cloudera Manager to encrypt sensitive information stored in the Cloudera Manager database by configuring a Credential Storage Provider (CSP).

Cloudera Manager stores a variety of sensitive information required for normal operations. This sensitive information is stored in plain text, either in the Cloudera Manager database or on disk.

You can configure Cloudera Manager to encrypt these sensitive values by configuring a Secure Credential Store that stores an encryption key to encrypt and decrypt sensitive information that are then stored in encrypted form only in the Cloudera Manager database. The following types of sensitive information can be encrypted:

- Configuration parameters containing usernames and passwords (except for those needed for Cloudera Manager to access the CSP).
- Kerberos keytabs

You can choose from the following types of CSP:

- Embedded – The credentials are stored on disk, on the Cloudera Manager server host that is protected by file permissions. This type is less secure than using a Vault, but is easier to set up and manage.
- Vault – You can install and configure an external Vault, located on a different host, if desired. Cloudera recommends using Vault from Hashicorp.

Known Limitations

There are currently the following limitations:

- Sensitive information that was written to the database before the CSP is enabled will not be encrypted automatically. If you change any sensitive information, it will be encrypted.

  You can regenerate Kerberos credentials, which will then be encrypted. To regenerate the credentials, go to Administration SecurityKerberos.
- Auto-TLS keys are not encrypted.
- The CSP Keystore Password, CSP Truststore Password and CM Truststore Password are not encrypted, as they are needed to connect to the CSP.
- The Cloudera Manager High Availability configuration is currently not supported with the Cloudera Manager Secure Credential Provider.

Configuring a Secure Credential Storage Provider for Cloudera Manager

Steps to configure Cloudera Manager to encrypt sensitive information stored in the Cloudera Manager database by configuring a Credential Storage Provider (CSP).

Before you begin

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

About this task



Important: You must complete the following steps, including uploading the certificates. If you fail to complete any of the following steps, then you cannot restart the Cloudera Manager server.

Procedure

1. If you are using the Vault, install and configure the Vault on a host with network connectivity to the Cloudera Manager server:
   a) Install and configure the Vault from Hashicorp.
   b) Configure TLS.
   c) Enable the Vault secrets engine at a path that will be dedicated to Cloudera Manager's use. Any Vault secrets engine is supported. See https://www.vaultproject.io/docs/secrets for more information. For example, the following enables a kv type secrets engine at path cm-secrets/:

      ```
      vault cm-secrets enable -path=secret kv
      ```

   d) A certificate authentication method must be configured at the path certs/cert1. This is how Cloudera Manager will authenticate with Vault. The method must have a policy that allows Create/Update/List/Read/Delete at the applicable secrets path. See https://www.vaultproject.io/docs/auth/cert for more information on certificate authentication and https://www.vaultproject.io/docs/concepts/policies for more information on vault policies.

      Example setup:

      ```
      vault auth enable cert 1
      vault write auth/cert/certs/cert1 display_name=example policies=example-
      policy certificate=@your-ca-cert.pem
      ```

      Example policy:

      ```
      path "cm-secrets/*"
      {
       capabilities = ["create", "read", "update", "list", "delete"]
      }
      ```

   e) The Vault must be unsealed by the user whenever Cloudera Manager is running. See https://www.vaultproject.io/docs/concepts/seal for instructions on unsealing Vault.
2. Configure the Secure Credential Store in Cloudera Manager:
   a) Open the Cloudera Manager Admin Console.
   b) Go to Administration and select the Security category.
   c) Select Enable Credential Storage to enable CSP. Under Credential Storage Provider Type, choose one of the following :

      - Embedded – The credentials are stored on disk, on the Cloudera Manager server host that is protected by file permissions. This type is less secure than using a Vault, but is easier to set up and manage.
      - Vault – You can install and configure an external Vault, located on a different host, if desired. Cloudera recommends using Vault from Hashicorp.

      If you select Embedded as Credential Storage Provider Type, then you may configure the following parameter if necessary:

      - Storage Path – Using this paramater you can configure the storage path for the file where you can store the encryption key.

      If you select Vault as Credential Storage Provider Type, then configure the following parameters:

- Vault Address – enter the URL of the host where you install the Vault.
- Secret Store Path – It specifies a path on the host where Vault stores the credentials. You cannot use this specified directory for any other purpose.

d) Click Save Changes.

3. Upload the following certificates, by copying them to the Cloudera Manager server host to the default directory location /opt/cloudera/csp-data, using the exact file name shown below, or you can use the Cloudera Manager API to upload the certificates using the CspResource endpoint. You can use the Cloudera Manager API Explorer to do this. Go to Support API Explorer .

Certificate	File name on the Cloudera Manager Host
Certificate authority	csp-ca.pem
Client certificate	csp-client.pem
Client key	csp-client.key

Important: The default location for storing the Vault certificates on the Cloudera Manager Server host is /opt/cloudera/csp-data.

Results

Sensitive information is now encrypted in the Cloudera Manager database.

Disabling the Credential Storage Provider

To disable the Credential Storage Provider (CSP), perform the following steps on the Cloudera Manager UI.

Before you begin

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

About this task

If you are already using a Vault based CSP and need to switch to a new Vault configuration, you need to disable CSP and enable CSP again with the new Vault configuration.

Procedure

1. Log in to the Cloudera Manager Admin Console.
2. Go to Administration and select the Security category.
3. To dissable CSP, click Disable Credential Storage and click OK.
4. If you want to enable a different type of CSP, see Configuring a Secure Credential Storage Provider for Cloudera Manager on page 140.

Results

The CSP is now disabled and the credentials and configurations stored in the Cloudera Manager are decrypted.

Managing Credential Storage Provider

After you enable Credential Storage Provider (CSP) successfully, you have the following options on Cloudera Manager UI to manage the CSP operations.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

Update Credential Storage Type

Use this option when you work on the following scenarios:

- When you want to update your Vault based CSP to an Embedded CSP
- When you want to update your Embedded CSP to a Vault based CSP

Important: To migrate from an existing Vault to a new Vault, you may need to move to an embedded CSP. Later, configure your new Vault and then update your CSP to use this new Vault after confirming its availability.

Key Rotation

To ensure consistent protection of your sensitive data from internal attacks, Cloudera recommends you to rotate encryption keys in a period of time (at least once every 4 to 5 years). This option decrypts the sensitive records and encrypts them again with a new encryption key and uses this key to encrypt any new records.

Disable Credential Storage

If you are already using a Vault based CSP and need to switch to a new Vault configuration, you need to disable CSP and enable CSP again with the new Vault configuration. This option decrypts all configurations encrypted through CSP and later disables CSP. By default, this option does not decommission your Vault server. Also, this option does not delete any older encryption keys saved in the CSP.

Resource Management

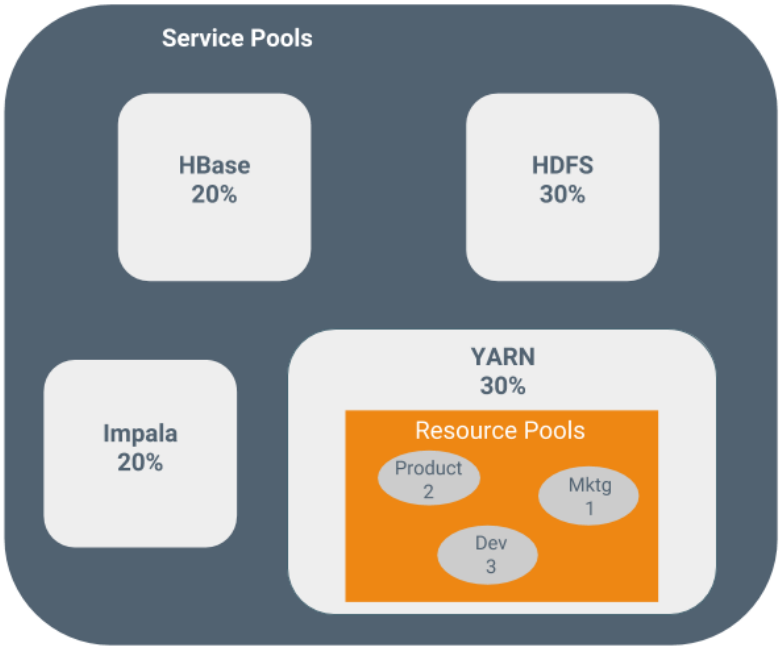
Resource management helps ensure predictable behavior by defining the impact of different services on cluster resources.

Use resource management to:

- Guarantee completion in a reasonable time frame for critical workloads.
- Support reasonable cluster scheduling between groups of users based on fair allocation of resources per group.
- Prevent a single user from creating a noisy neighbor situation that deprives other users access to the cluster.

Statically allocating resources using Control Groups (cgroups) is configurable through a single static service pool wizard. You allocate services as a percentage of total resources, and the wizard configures the cgroups.

For example, the following figure illustrates static pools for HBase, HDFS, Impala, and YARN services that are respectively assigned 20%, 30%, 20%, and 30% of cluster resources.



You can dynamically apportion resources that are statically allocated to YARN and Impala by using dynamic resource pools.

Depending on the version of Cloudera Runtime you are using, dynamic resource pools in Cloudera Manager support the following scenarios:

- YARN - YARN manages the virtual cores, memory, running applications, maximum resources for undeclared children (for parent pools), and scheduling policy for each pool. In the preceding diagram, three dynamic resource pools—Dev, Product, and Mktg with weights 3, 2, and 1 respectively—are defined for YARN. If an application starts and is assigned to the Product pool, and other applications are using the Dev and Mktg pools, the Product resource pool receives 30% x 2/6 (or 10%) of the total cluster resources. If no applications are using the Dev and Mktg pools, the YARN Product pool is allocated 30% of the cluster resources.
- Impala - Impala manages memory for pools running queries and limits the number of running and queued queries in each pool.

Static Service Pools

Static service pools is a Cloudera Manager feature that allows you to partition cluster resources by service.

This feature provides a simplified, wizard-driven workflow for configuring the underlying Linux Control Groups (cgroups) on your cluster hosts. By allocating a static percentage of resources—CPU, memory, and I/O—to each service, you can prevent a high-load service from consuming all host resources and negatively impacting other services, a situation often referred to as the *noisy neighbor* problem.

Resource Management is implemented at the role group level within a cluster. When you complete the wizard, Cloudera Manager automatically calculates and applies the appropriate cgroup parameters and application-level memory configurations based on the percentages you have assigned.

The Static Service Pools wizard accomplishes two primary configuration tasks:

- **Cgroup Parameter Configuration:** It translates your percentage-based allocations into specific Linux cgroup parameter values for each service's roles. For example, it sets CPU shares/weights and I/O weights to manage resource contention.
- **Cooperative Memory Management:** Instead of relying solely on the cgroup memory soft and hard limits, which can cause the kernel to terminate processes abruptly, Static Service Pools also set *cooperative* memory limits.

This is primarily achieved by configuring the maximum Java heap size for Java-based services. This ensures that applications are aware of their memory boundaries and can handle memory allocation failures more gracefully, providing a better user experience than having processes unexpectedly killed by the kernel's Out-Of-Memory (OOM) killer.

It is important to note that the behavior of resource control, particularly for I/O, differs between cgroup v1 and cgroup v2. Cgroup v1 I/O weighting primarily affects read requests, whereas cgroup v2 provides more comprehensive control over both read and write I/O. The Static Service Pool wizard presents distinct configuration tables for cgroup v1 and cgroup v2 to account for these differences.

Note:

- I/O allocation only works when short-circuit reads are enabled.
- I/O allocation does not handle write-side I/O because cgroups in the Linux kernel do not currently support buffered writes.

Viewing Static Service Pool Status

Select Clusters *CLUSTER NAME* Static Service Pools . If the cluster has a YARN service, the Static Service Pools Status tab displays and shows whether resource management is enabled for the cluster and the currently configured service pools.

Frequently Asked Questions

Which worker roles get configured by Static Service Pools?

The following worker roles are configured by the Static Service Pools wizard for each service:

- Impala Service: Impala Daemon role
- Solr Service: Solr Server role
- YARN Service: Nodemanager role
- HDFS Service: Datanode role
- HBase Service: RegionServer role
- MapReduce Service: TaskTracker role

Are Linux Control Groups (cgroups) used for managing memory limits?

No, memory management is controlled through Java or service-specific configurations.

Why can’t I use Linux Control Groups (cgroups) soft and hard limits for managing memory?

When a service exceeds a cgroup hard limit, the Linux kernel might abruptly terminate the entire process to enforce the limit. This is a harsh action that can cause unexpected service outages and is handled by the kernel's Out-Of-Memory (OOM) killer. While memory soft limits are less aggressive than hard limits, they are avoided for the same fundamental risk that the Linux kernel will unexpectedly terminate a service process.

How do the weights assigned to each service for CPU and I/O in Linux Control Groups compare to the default weights assigned to other processes running on the system?

For CGroups V2, the default weight assigned to a process is 100. Therefore, Cloudera processes assigned a weight higher than 100 will gain a greater share of CPU cycles or I/O bandwidth than other system processes. For CGroups V1, the default is 1024, but the same concept of relative proportions applies to the number of resources allocated to a process.

Why do I see a value of -1 for a Linux Control Group (cgroup) configuration?

This is the default Cloudera Manager value put in place with Linux Control Groups are not enabled.

Does using the Static Service Pools wizard affect the amount of memory allocated to containers created by YARN?

No, the memory settings configured by the Static Service Pools wizard only apply to the NodeManager role's process. The amount of memory available for YARN containers is determined by the yarn.nodemanager.resource.memory-mb property, which is assigned a default value by Cloudera Manager.

Can I use CGroup configurations to limit how much memory YARN containers can use?

You should use the memory setting yarn.nodemanager.resource.memory-mb to set how much memory YARN can allocate to containers it creates. The setting yarn.nodemanager.resource.memory.cgroups.soft-limit-percentage might be confused with such control, but is actually used to throttle memory utilization within a container once it reaches the set limit, which has a default value of 90%.

Enabling and Configuring Static Service Pools

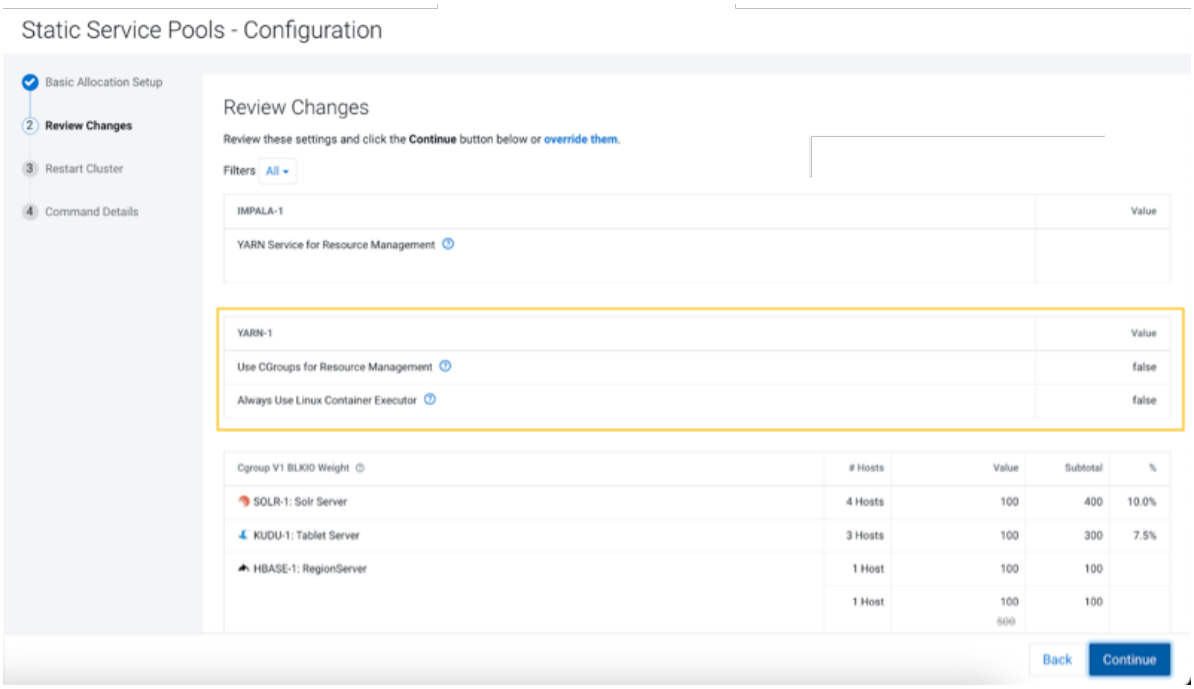
To enable and configure static service pools, you enter the percentage of resources to allocate to each service and then restart the cluster.

Before you begin

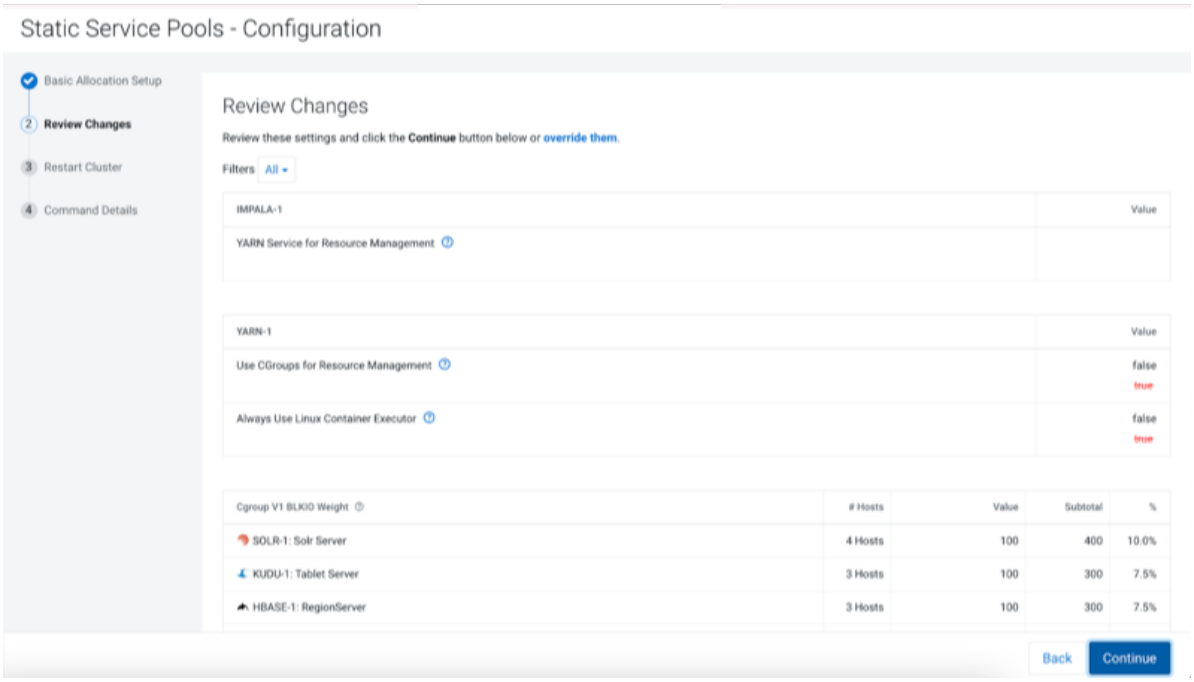
Minimum Required Role: Cluster Administrator (also provided by Full Administrator)

Procedure

1. Select Clusters*CLUSTER NAME*Static Service Pools.
2. Click the Configuration tab.
   The Step 1 of 4: Basic Allocation Setup page displays. In each field in the basic allocation table, enter the percentage of resources to give to each service. The total must add up to 100%.
3. After allocating shares to services in the Static Service Pools wizard in step 2, the user is redirected to the Review Allocations page. Static Service Pool always sets the following YARN's configurations Use Cgroups for Resource Management and Always Use Linux Container Executor to FALSE regardless if these configurations were set to TRUE.
   a. If these two configurationss were set to FALSE before running the wizard, then Static Service Pool does not set them to TRUE. See the following figure:

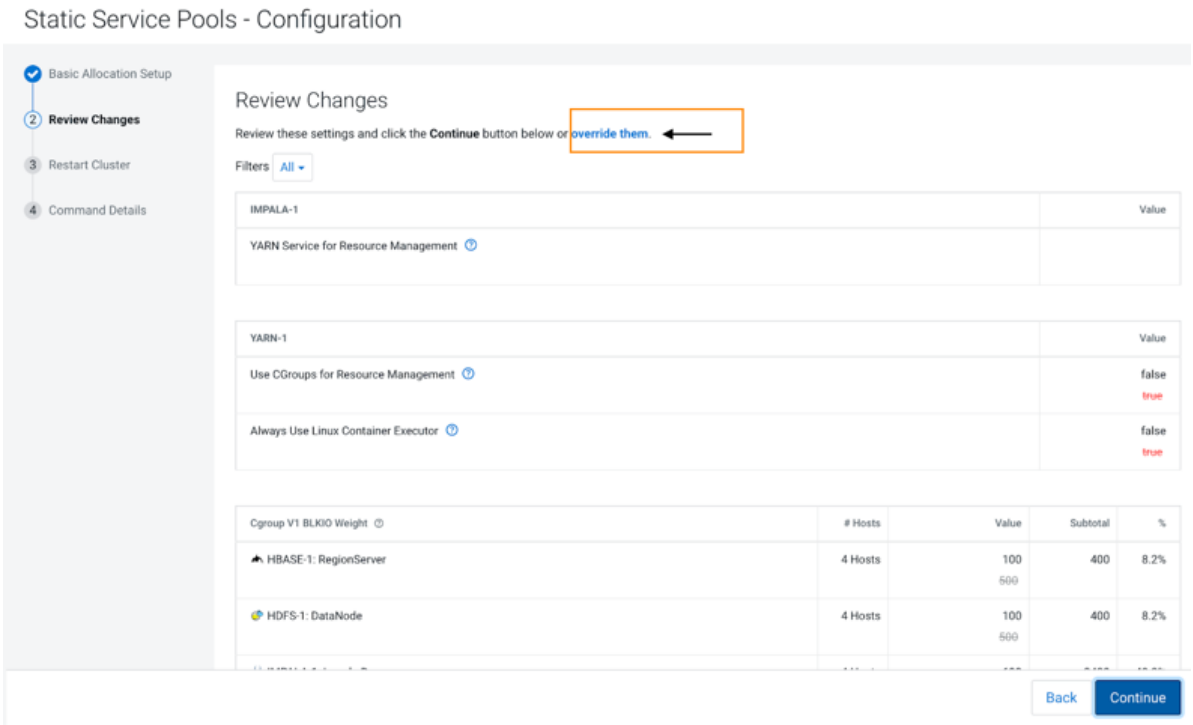


b. If these two configurations were set to TRUE, then Static Service Pool explicitly sets them to FALSE. This is a bug. See the following figure:

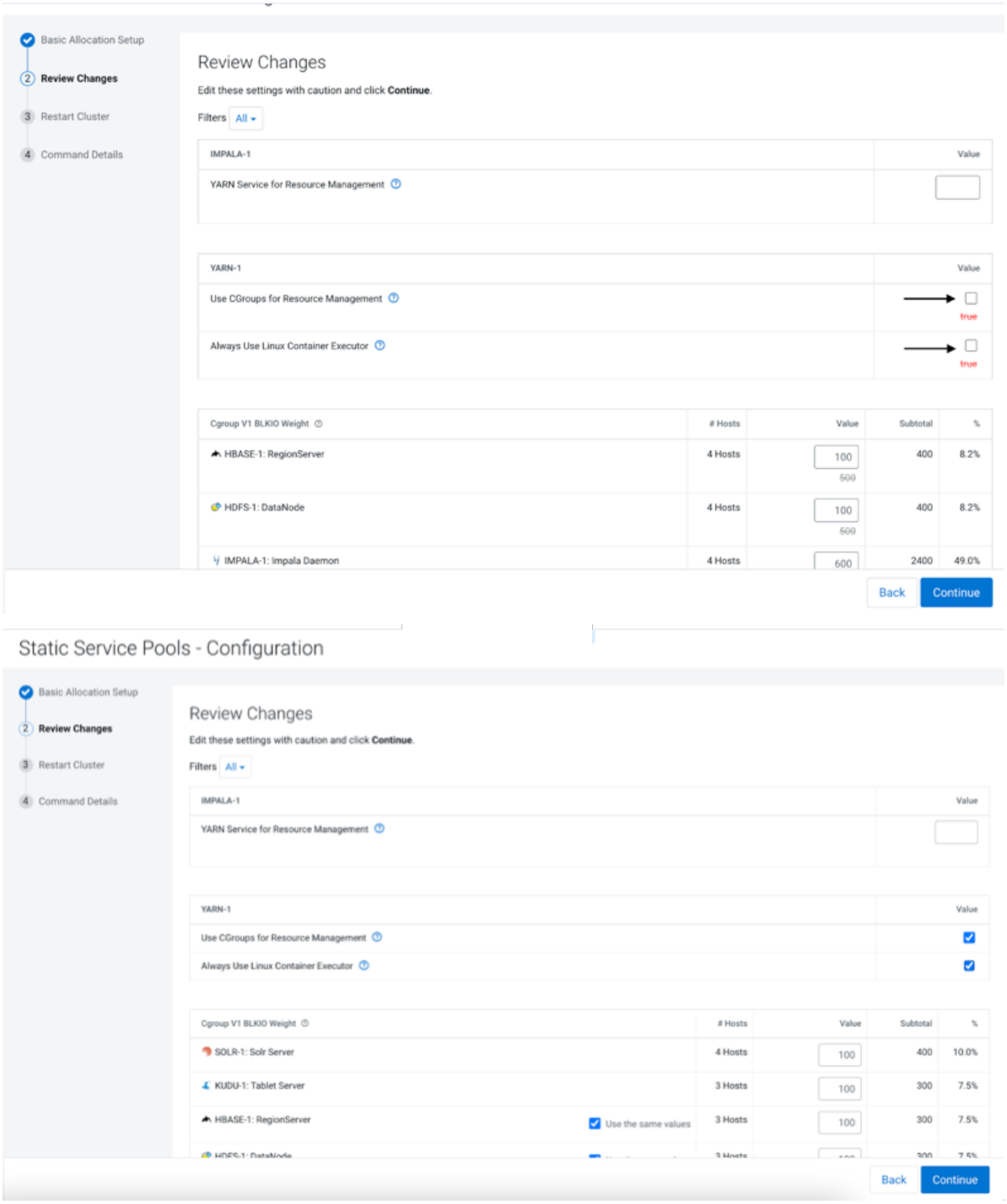


c. You must enable these two configurations for Cgroup v2.

d. You can enable these two configurations by selecting the override them button on the page.



e. Then select those YARN configurations check boxes to set to TRUE or keep them set as TRUE. See the following figure:



Static Service Pools - Configuration

4. Click Continue to proceed.
   Step 2 of 4: Review Changes - The allocation of resources for each resource type and role displays with the new values as well as the values previously in effect. The values for each role are set by role group; if there is more than one role group for a given role type (for example, for RegionServers or DataNodes) then resources will be allocated separately for the hosts in each role group.

   Important:

   For each service there are separate resource allocation tables for cgroup v1 and cgroup v2 which display percentage allocations for different resource configuration parameters.

5. Take note of changed settings. If you have previously customized these settings, then check the resource allocations for each service based on your cgroup version carefully. You can change your resource allocations by clicking the Back button to go to the previous page and change your allocations.
6. When you are satisfied with the allocations, click Continue.
   The Step 3 of 4: Restart Services page displays.
7. To apply the new allocation percentages, click Restart Now to restart the cluster. To skip this step, click Restart Later. If HDFS High Availability is enabled, you will have the option to choose a rolling restart.
8. Step 4 of 4: Progress displays the status of the restart commands. Click Finished after the restart commands complete.

After you enable static service pools, there are three additional tasks:

9. Delete everything under the local directory path on NodeManager hosts. The local directory path is configurable, and can be verified in Cloudera Manager with YARN Configuration NodeManager Local Directories .
10. Enable cgroups for resource management. You can enable cgroups in Cloudera Manager with Yarn Configuration Use CGroups for Resource Management .
11. If you are using the optional Impala scratch directory, delete all files in the Impala scratch directory. The directory path is configurable, and can be verified in Cloudera Manager with Impala Configuration Impala Daemon Scratch Directories .

Describing different columns in a Static Service Pool Wizard

Static Service Pool (SSP) resource percentages provided in the Cloudera Manager Wizard are used as inputs to calculate internal control values such as cpu.shares, cpu.weight, blkio.weight, and io.weight. These values are not expressed as percentages in the system but as raw numbers used by the Linux cgroups subsystem to allocate resources relatively among Runtime services.

How Weights Are Calculated

The following formulas are used to compute actual values based on percentages input in the wizard:

Resource Type	Formula	Minimum value
cpu.shares	max (20, percentage * 20)	20
cpu.weight	max(2, percentage * 2)	2
blkio.weight	max(100, percentage * 10)	100
io.weight	max(20, percentage * 2)	20



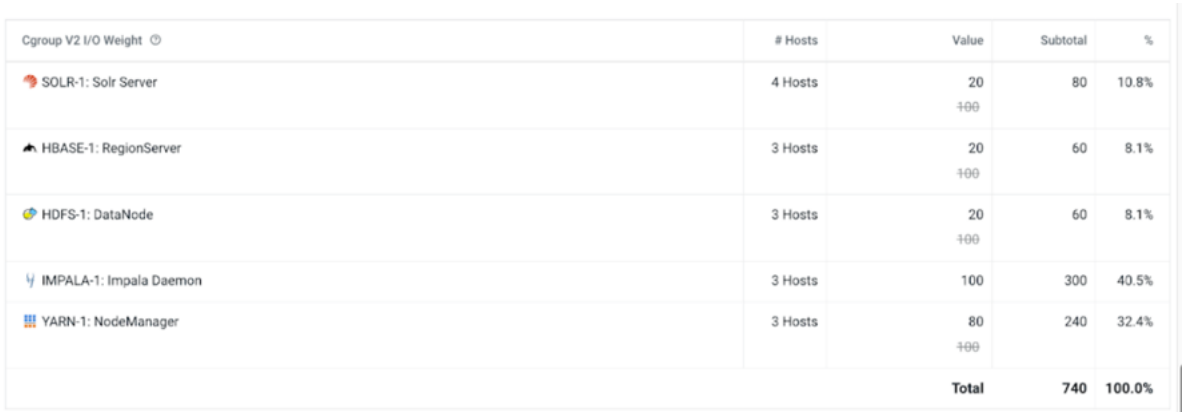
Important:

For example, if you assign **50%** to Impala:

```
io.weight = max(20, 50 * 2) = 100
```

Why do actual percentages on hosts differ?

The percentages visible in the *Resource Usage tables*, such as for io.weight, are relative values derived from **actual** weight * number of hosts used by each role group.



Cgroup V2 I/O Weight	# Hosts	Value	Subtotal	%
SOLR-1: Solr Server	4 Hosts	20 ~~100~~	80	10.8%
HBASE-1: RegionServer	3 Hosts	20 ~~100~~	60	8.1%
HDFS-1: DataNode	3 Hosts	20 ~~100~~	60	8.1%
IMPALA-1: Impala Daemon	3 Hosts	100	300	40.5%
YARN-1: NodeManager	3 Hosts	80 ~~100~~	240	32.4%
		Total	**740**	**100.0%**

Role Group	Hosts	io.weight (Value)	Subtotal = Hosts × Value	% of Total
Impala Daemon	3	100	300	300 / 740 = 40.5%
YARN NodeManager	3	80	240	240 / 740 = 32.4%
HDFS DataNode	3	20	60	60 / 740 = 8.1%
Other Services such as Solr and HBase	—	—	140	140 / 740 = 18.9%

So, while you might have **assigned 50% to Impala** in the wizard, its **actual share may appear as ~40%** depending on the following parameters:

- Number of hosts per service
- Actual calculated value per host
- Presence of other Runtime services such as Solr, Kudu, and HBase which also receive minimum weights even if **not assigned** in the wizard

> **Important:** If other Runtime services like Solr, Kudu, and HBase are active but not explicitly assigned percentages in the wizard, they will still:
>
> - Be part of resource competition
> - Be assigned **minimum resource weights** through fallback logic

Disabling Static Service Pools

To disable static service pools, disable cgroup-based resource management for all hosts in all clusters.

Before you begin

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

Procedure

1. In the main navigation bar, click Hosts.
2. Click the Configuration tab.
3. Select ScopeResource Management.
4. Clear the Enable Cgroup-based Resource Management property.
5. Click Save Changes.
6. Restart all services.

Results

Static resource management is disabled, but the percentages you set when you configured the pools, and all the changed settings (for example, heap sizes), are retained by the services. The percentages and settings will also be used when you re-enable static service pools. If you want to revert to the settings you had before static service pools were enabled, follow the procedures in *Viewing and Reverting Configuration Changes*.

Linux Control Groups (cgroups)

Administrators can use Cloudera Manager to configure the Linux Control Groups (cgroup v1 or cgroup v2) kernel feature to manage host resources, such as CPU and I/O, to prevent services from negatively impacting each other. This situation, often called the *noisy neighbor* problem, can overcome by administrators by imposing per-resource restrictions on services and roles to isolate compute frameworks from one another.

Minimum Required Role: Full Administrator. This feature is not available when using to manage Data Hub clusters.

You can configure these resource allocations in two ways:

- **Static Service Pools:** This is the recommended, wizard-driven method that simplifies cgroup configuration by allowing you to allocate resources as percentages.
- **Manual Configuration:** For advanced control, you can set individual cgroup properties directly on cluster services and roles.



Important: Cloudera does not recommend using cgroups soft or hard limits for memory management.

While cgroups can manage memory, Cloudera Manager primarily uses them for CPU and I/O contention. For memory, Cloudera Manager sets application-level limits (For example, the maximum Java heap size). This approach is designed to prevent the Linux kernel's Out-Of-Memory (OOM) killer from abruptly terminating a service process, which can occur when a cgroup memory hard limit is breached. While memory soft limits are less aggressive than hard limits, they are avoided for the same fundamental risk that the Linux kernel will unexpectedly terminate a service process.

In case of cgroup v1, if you have configured cgroups in the Linux environment for your cluster hosts, you can choose to use those custom cgroups instead of the default cgroups configurations provided by Cloudera Manager.



Attention: You cannot mix these Cloudera Manager-managed cgroups with custom cgroups that are managed in your Linux environment.

Enabling Resource Management with Control Groups

Enabling Linux Control Groups (cgroups) using Cloudera Manager.

About this task

You can configure Cloudera Manager to use Linux Control Groups (cgroups) to manage cluster resources. After enabling cgroups for Resource management, you use Service configuration properties to allocate resources by CPU shares, I/O, and memory.

About this task

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

Cgroups-based resource management can be enabled for all hosts, or on a per-host basis.

Procedure

1. Click Hosts Host Configuration.
2. Click CategoryResource Management.
3. Select the Enable Cgroup-based Resource Management parameter.

4. To enable Cgroups only on specific hosts:
   a) Click the Add Host Overrides link.
      The Add Host Overrides page displays.
   b) Select the hosts where you want to enable Cgroups.
   c) Click the Add button.
      The Host Configuration page displays.
   d) De-select the All Hosts option in the Enable Cgroup-based Resource Management configuration.
   e) Click Save Changes.
5. Restart all roles on the host(s).
6. To configure the default Cloudera Manager resource parameters, see Configuring Resource Parameters on page 152. If you are using Custom Cgroups to allocate resources, you configure those resource parameters in the Linux environment.

Limitations

- Role group and role instance override cgroup-based resource management parameters must be saved one at a time. Otherwise some of the changes that should be reflected dynamically will be ignored.
- The role group abstraction is an imperfect fit for resource management parameters, where the goal is often to take a numeric value for a host resource and distribute it amongst running roles. The role group represents a "horizontal" slice: the same role across a set of hosts. However, the cluster is often viewed in terms of "vertical" slices, each being a combination of worker roles (such as TaskTracker, DataNode, RegionServer, Impala Daemon, and so on). Nothing in Cloudera Manager guarantees that these disparate horizontal slices are "aligned" (meaning, that the role assignment is identical across hosts). If they are unaligned, some of the role group values will be incorrect on unaligned hosts. For example a host whose role groups have been configured with memory limits but that's missing a role will probably have unassigned memory.

Configuring Resource Parameters

After enabling cgroups, you can restrict and limit the resource consumption of roles (or role groups) on a per-resource basis by using resource allocation configuration parameters.

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

Before you begin:

OPSAPS-74341: NodeManagers might fail to start during the cluster restart after the Cloudera Manager 7.13.1.x upgrade

Cgroup v2 support is enabled in CDP 7.1.9 SP1 CHF5 and higher versions. However, if the user upgrades from Cloudera Manager 7.11.3.x to Cloudera Manager 7.13.1.x, and the environment is using cgroup v2, the NodeManagers might fail to start during the cluster restart after the Cloudera Manager 7.13.1.x upgrade.

Workaround:

To resolve this issue temporarily, you must perform the following steps:

1. Go to the YARN service page on the Cloudera Manager UI.
2. Navigate to the Configuration tab.
3. Search for NodeManager Advanced Configuration Snippet (Safety Valve) for yarn-site.xml.
4. Add the following entry:
   a. Add yarn.nodemanager.linux-container-executor.cgroups.v2.enabled=true
5. Restart the Nodemanagers. Nodemanagers restart successfully.

Steps:

You can find all of these resource parameters in the Cloudera Manager Admin Console, under the Resource Management category. To change resource allocations:

1. In the Cloudera Manager Admin Console, go to the service where you want to configure resources.

2. Click the Configuration tab.
3. Select Category Resource Management .
4. Locate the resource allocation configuration parameters that start with Cgroup V1 (if you are using cgroup v1) or Cgroup V2 (if you are using cgroup v2). You can specify resource allocations for each type of resource (CPU, memory, etc.) for all roles of the service using these parameters, or you can specify different allocations for each role by clicking the Edit Individual Values link. Edit any of the following parameters:

   - **Cgroup V1 parameters:**
     - Cgroup V1 CPU Shares - The more CPU shares given to a role, the larger its share of the CPU when under contention. Until processes on the host (including both roles managed by Cloudera Manager and other system processes) are contending for all of the CPUs, this will have no effect. When there is contention, those processes with higher CPU shares will be given more CPU time. The effect is linear: a process with 4 CPU shares will be given roughly twice as much CPU time as a process with 2 CPU shares.
     - Cgroup V1 BLKIO Weight - The greater the I/O weight, the higher priority will be given to I/O requests made by the role when I/O is under contention (either by roles managed by Cloudera Manager or by other system processes).

       This only affects read requests; write requests remain unprioritized. The Linux I/O scheduler controls when buffered writes are flushed to disk, based on time and quantity thresholds. It continually flushes buffered writes from multiple sources, not certain prioritized processes.
     - Cgroup V1 Memory Soft Limit - When the limit is reached, the kernel will reclaim pages charged to the process if and only if the host is facing memory pressure. If reclaiming fails, the kernel may stop the process. Both anonymous as well as page cache pages contribute to the limit.
     - Cgroup V1 Memory Hard Limit - When a role's resident set size (RSS) exceeds the value of this parameter, the kernel will swap out some of the role's memory. If it is unable to do so, it will stop the process. The kernel measures memory consumption in a manner that does not necessarily match what the top or ps report for RSS, so expect that this limit is a rough approximation.
   - **Cgroup V2 parameters:**
     - Cgroup V2 CPU Weight - The more CPU weight given to a role in cgroup V2, the larger its share of the CPU when under contention. Until processes on the host (including both roles managed by Cloudera Manager and other system processes) are contending for all of the CPUs, this setting will have no effect. When there is contention, those processes with higher CPU weights will be given more CPU time. The allocation is proportional: a role with a CPU weight of 200 will receive twice as much CPU time as a role with a weight of 100.
     - Cgroup V2 I/O Weight - The greater the I/O weight assigned to a role in cgroup V2, the higher priority its I/O requests will receive when there is contention for I/O resources. This means that roles with higher I/O weights will have their read and write requests processed more frequently compared to roles with lower weights, improving their overall performance under I/O contention.

       Unlike cgroup V1, cgroup V2 prioritizes both read and write requests, allowing for a more balanced resource allocation. The impact is proportional: a role with an I/O weight of 200 will have its requests prioritized over a role with a weight of 100.
     - Cgroup V2 Memory Soft Limit - In cgroup V2, the memory soft limit allows a role to use memory up to a specified limit, with leniency under normal conditions. If the host experiences memory pressure, the kernel will preferentially reclaim memory from processes that exceed their soft limits, attempting to free up space. If reclamation is unsuccessful and memory pressure remains, then the kernel might kill processes to alleviate the situation. Both anonymous and page cache pages count towards this limit, influencing which memory is reclaimed. Unlike hard limits, this approach provides flexibility under low memory pressure, while still protecting system stability during high memory usage scenarios.
     - Cgroup V2 Memory Hard Limit - In cgroup V2, the memory hard limit strictly confines a role's memory usage to the specified value. When the memory consumption of a role, including its Resident Set Size (RSS), exceeds this limit, the kernel will attempt to swap out some of the role's memory to stay within bounds. If swapping is not sufficient or feasible, the kernel will terminate the process to enforce the limit. Memory consumption is tracked more comprehensively than what utilities like top or ps report for RSS, so the hard limit serves as a strict boundary rather than an exact match to these metrics.

5. Click Save Changes.
6. Restart the service or roles where you changed values.

   See Restarting a Cloudera Runtime Service or Starting, Stopping, and Restarting Role Instances on page 61.

Protecting Production MapReduce Jobs from Impala Queries

Suppose you have MapReduce deployed in production and want to roll out Impala without affecting production MapReduce jobs. For simplicity, we will make the following assumptions:

- The cluster is using homogenous hardware
- Each worker host has two cores
- Each worker host has 8 GB of RAM
- Each worker host is running a DataNode, TaskTracker, and an Impala Daemon
- Each role type is in a single role group
- Cgroups-based resource management has been enabled on all hosts

Action	Procedure
CPU	1. Leave DataNode and TaskTracker role group CPU shares at 1024. 2. Set the Impala Daemon role group's CPU shares to 256. 3. The TaskTracker role group should be configured with a Maximum Number of Simultaneous Map Tasks of 2 and a Maximum Number of Simultaneous Reduce Tasks of 1. This yields an upper bound of three MapReduce tasks at any given time; this is an important detail for memory sizing.
Memory	1. Set the Impala Daemon role group memory limit to 1024 MB. 2. Leave DataNode maximum Java heap size at 1 GB. 3. Leave TaskTracker maximum Java heap size at 1 GB. 4. Leave MapReduce Child Java Maximum Heap Size for Gateway at 1 GB. 5. Leave cgroups hard memory limits alone. We'll rely on "cooperative" memory limits exclusively, as they yield a nicer user experience than the cgroups-based hard memory limits.
I/O	1. Leave DataNode and TaskTracker role group I/O weight at 500. 2. Impala Daemon role group I/O weight is set to 125.

When you're done with configuration, restart all services for these changes to take effect. The results are:

1. When MapReduce jobs are running, all Impala queries together will consume up to a fifth of the cluster's CPU resources.
2. Individual Impala Daemons will not consume more than 1 GB of RAM. If this figure is exceeded, new queries will be cancelled.
3. DataNodes and TaskTrackers can consume up to 1 GB of RAM each.
4. We expect up to 3 MapReduce tasks at a given time, each with a maximum heap size of 1 GB of RAM. That's up to 3 GB for MapReduce tasks.
5. The remainder of each host's available RAM (6 GB) is reserved for other host processes.
6. When MapReduce jobs are running, read requests issued by Impala queries will receive a fifth of the priority of either HDFS read requests or MapReduce read requests.

Operating System Support for Cgroups

Cgroups are a feature of the Linux kernel, and as such, support depends on the host's Linux distribution and version as shown in the following tables. If a distribution lacks support for a given parameter, changes to the parameter have no effect.

The exact level of support can be found in the Cloudera Manager Agent log file, shortly after the Agent has started. In the log file, look for an entry like this:

```
Found cgroups capabilities: {
'has_blkio': False, 'has_io': True, 'has_cpu': True, 'has_cpuacct': False,
'has_memory': True, 'has_devices': False, 'default_blkio_weight': -1,
'default_io_weight': -1, 'default_cpu_shares': -1, 'default_cpu_weight': -1
,
```

```
'default_cpu_max': -1, 'default_cpu_rt_runtime_us': -1, 'default_memory_l
imit_in_bytes': -1,
'default_memory_max': -1, 'default_memory_soft_limit_in_bytes': -1,
'default_memory_high': -1, 'writable_cgroup_dot_procs': False}
```

The has_cpu and similar entries correspond directly to support for the CPU, I/O, and memory parameters.

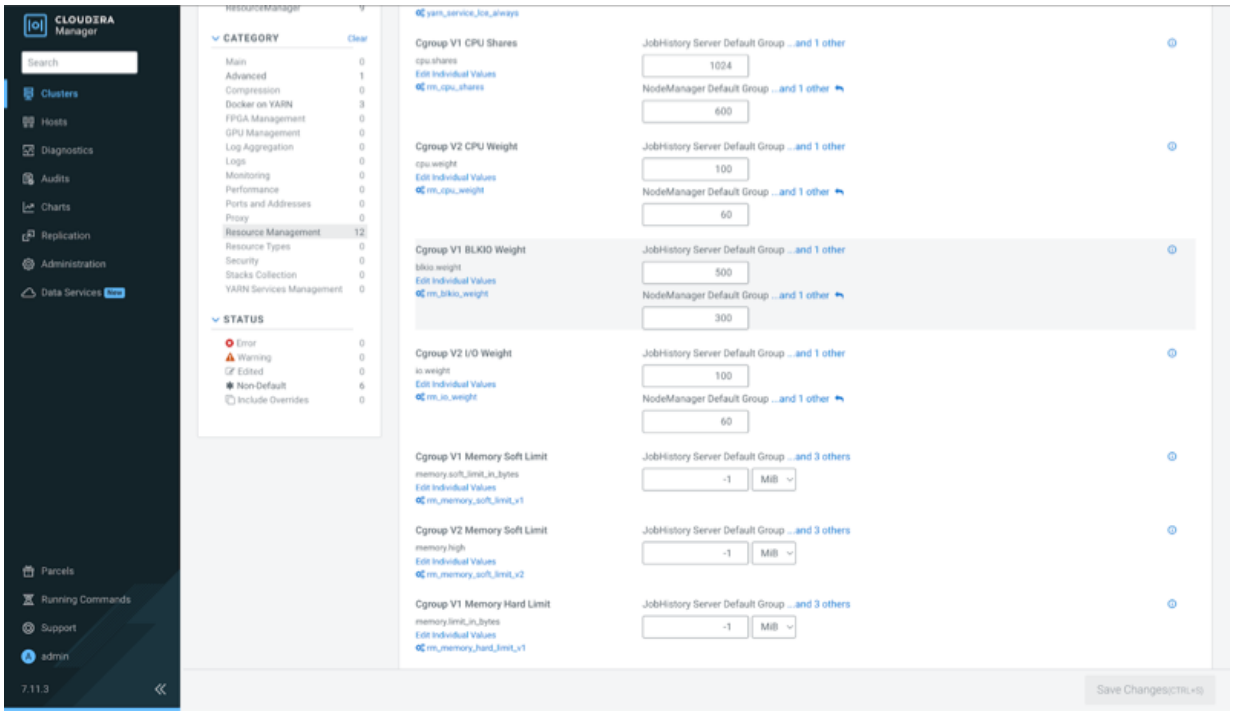
Verifying Cgroup Configuration Changes

When managing services through Cloudera Manager 7.13.1.x, any cgroup-related configuration changes made on a service's Configuration page are applied instantly and directly to the corresponding controller files on the relevant hosts. This allows for immediate verification of changes without requiring a service restart or waiting for stale configuration indicators.

To confirm that your cgroup configurations have taken effect:

1. Locate the Configuration: Navigate to the Configuration page for your desired service such as YARN.
2. Modify a Cgroup Setting: Update any Cgroup-related parameter such as Cgroup V2 CPU Weight for JobHistory roles as shown in the following figure.
3. Instant Reflection: The changes are immediately written to the respective Cgroup controller files on the host where the role is running.

Figure 3: Verifying YARN JobHistory for Cgroup V2 CPU Weight



Suppose you are viewing the Cgroup configurations on YARN's configuration page. If you modify the cpu.weight value:

Any changes you make in Cloudera Manager to Cgroup V2 CPU Weight for this role will instantly update this file, which you can verify with the following `cat` command:

```
# SSH into the relevant host where the JobHistory role is running # Navi
gate to the cGroup directory for the JobHistory service
[root@bofa-testcgroups1-1 cgroup]# cd /sys/fs/cgroup/330-yarn-JOBHISTORY

# View the current cpu.weight value in the controller file
[root@bofa-testcgroups1-1 330-yarn-JOBHISTORY]# cat cpu.weight
100
```

Important: Cgroup V1 vs. Cgroup V2

The visibility of Cgroup configuration changes depends on the Cgroup version enabled on your system:

- If Cgroup V2 is enabled, you will only see changes reflected in Cgroup V2 controller files (such as Cgroup V2 CPU Weight as shown in this example. Cgroup V1 controller files will not be present or updated, even if you configure Cgroup V1 settings in Cloudera Manager UI.
- If Cgroup V1 is enabled (default on older systems), changes to Cgroup V1 configurations will be visible in the Cgroup V1 controller files.

Always ensure you are checking the controller files corresponding to the Cgroup version active on your operating system.

Configuring Custom Cgroups

Instead of using the default Cloudera Manager cgroups for resource management, you can configure custom cgroups only in the case of cgroup v1. You must configure these cgroups in the Linux environments of your cluster hosts before enabling custom cgroups in Cloudera Manager. You can configure Custom cgroups for all or selected roles of a service.

About this task

Important:

You can configure custom cgroups only in case of cgroup v1.

When you configure a custom cgroup for any service or role, all of the other cgroup configurations managed by Cloudera Manager are ignored. You cannot choose to only override a single cgroup subsystem. Any subsystems not defined in the custom cgroup configuration parameter will be set to the operating system default.

Procedure

1. If you have not already enabled cgroups, enable Resource Management with Control Groups. See Enabling Resource Management with Control Groups on page 151.
2. In the Cloudera Manager Admin Console, go to the service where you want to enable custom cgroups.
3. Click the Configuration tab.
4. Select CategoryResource Management.
5. Locate the Custom Control Group Resources (overrides Cgroup settings) configuration parameter.
6. To configure cgroups for all roles of the service:
   a) Enter the names of the pre-configured cgroups you want to apply to the service. Separate the cgroup names using a single space. Use the following format:

   ```
   <SUBSYSTEM>:<PATH>
   ```

   or

   ```
   <SUBSYSTEM>,<SUBSYSTEM>, ...:<PATH>
   ```

   This is the same format used to launch processes from the Linux command line using the cgexec command.

   For example, to enable a CPU control group:

   ```
   cpu:myCgroup
   ```

   For more information, see the documentation for your operating system.
   b) Click Save Changes.
   c) Restart the service. See Restarting a Cloudera Runtime Service.

7. To enable a custom cgroup for one or more roles of this service:
   a) Click the Edit Individual Values link.
      A text box for each role of this service displays.
   b) Enter the names of the pre-configured cgroups you want to apply to each role using the format described in the previous step.
   c) Restart each role where you configured a custom cgroup. See Starting, Stopping, and Restarting Role Instances on page 61.

Migrating from Cgroup v1 to Cgroup v2

Learn how to migrate from Cgroup v1 to Cgroup v2. You must migrate from Cgroup v1 to Cgroup v2 for managing the cluster resources using Cgroup v2 resource allocation configuration parameters.

About this task



Important:

- Cloudera does not support Cgroup v2 on RHEL 7, as it is a deprecated operating system. Therefore, no support or testing is provided for cgroups v2 on the RHEL 7 platform.
- Cloudera Manager currently does not support Cgroup v2 on Ubuntu 22.04. Additionally, Cloudera Manager does not support hybrid Cgroup configurations where both Cgroup v1 and v2 coexist.

Procedure

1. Upgrade Cloudera Manager to Cloudera Manager 7.13.1.500 CHF5 (minimum Cloudera Manager version which supports Cgroup v2). For upgrading Cloudera Manager, see Upgrading Cloudera Manager 7.

   Important:

   For the users using RHEL 9.x with Cloudera Manager version lower than 7.13.1.500 CHF5, must disable Cgroup v2 if already enabled before upgrading to Cloudera Manager 7.13.1.500 CHF5 version as Cgroup v2 is not supported with Cloudera Manager version lower than 7.13.1.500 CHF5.
2. If you want to provide a custom path in the config.ini file for Cgroup V2, ensure you add the following parameter to the config.ini file under the [Cgroup_Paths] section after performing a Cloudera Manager upgrade:

   ```
   controllers_cgroup_path=/
   ```

   In case of fresh installation, by default, this custom path is available in the config.ini file under the [Cgroup_Paths] section and hence there is no need to add a custom path in the config.ini file for cgroup V2.
3. Upgrade CDP Runtime to CDP 7.3.1 (minimum CDP Runtime version). For upgrading CDP Runtime, see Upgrading a Cloudera Runtime Cluster.
4. Disable the Enable Cgroup-based Resource Management property for Cgroup v1 by performing the following steps:
   a. On the Cloudera Manager UI, navigate to Hosts Hosts Configuration .
   b. Select Category Resource Management .
   c. Clear the Enable Cgroup-based Resource Management checkbox.
   d. Click Save Changes.
5. Upgrade OS to RHEL 9.

   Important:

   During major OS upgrades, while upgrading from Redhat 8 (defaults to cgroup v1) to Redhat 9 (defaults to Cgroup v2), the resource configurations will not be automatically transferred such as value of Cgroup V1 CPU Shares will not be populated in Cgroup V2 CPU Weight. Also, the controller files inside the process directories will be created under cgroups root path with default values.
6. Enable the Enable Cgroup-based Resource Management property for Cgroup v2 by performing the following steps:

a. On the Cloudera Manager UI, navigate to Hosts Hosts Configuration .
b. Select Category Resource Management .
c. Select the Enable Cgroup-based Resource Management checkbox.
d. Click Save Changes.



Important:

- If you are setting Cgroup v1 parameter values manually, then you should now set Cgroup v2 parameter values manually (performing conversion of values manually) and restart the services using cgroups.

  Note that Cloudera Manager UI will have old values under Cgroup v1 parameters which you can use as a reference to re-configure the values in the case of Cgroup v2.
- If you are using the **Static Service Pools** wizard to allocate resources, then you should run the **Static Service Pools** wizard again with the necessary percentage settings to generate Cgroup v2 parameter values.

7. Restart the services that are using cgroups.

Data Storage for Monitoring Data

The Service Monitor and Host Monitor roles in the Cloudera Management Service store time series data, health data, and Impala query and YARN application metadata.

Configuring Service Monitor Data Storage

The Service Monitor stores time series data and health data, Impala query metadata, and YARN application metadata.

By default, the data is stored in /var/lib/cloudera-service-monitor/ on the Service Monitor host. You can change this by modifying the Service Monitor Storage Directory configuration (firehose.storage.base.directory). To change this configuration on an active system, see *Moving Monitoring Data on an Active Cluster*.

You can control how much disk space to reserve for the different classes of data the Service Monitor stores by changing the following configuration options:

- Time-series metrics and health data - Time-Series Storage (firehose_time_series_storage_bytes - 10 GB default, 10 GB minimum)
- Impala query metadata - Impala Storage (firehose_impala_storage_bytes - 1 GB default)
- YARN application metadata - YARN Storage (firehose_yarn_storage_bytes - 1 GB default)

For information about how metric data is stored in Cloudera Manager and how storage limits impact data retention, see Data Granularity and Time-Series Metric Data.

The default values are small, so you should examine disk usage after several days of activity to determine how much space is needed.

Configuring Host Monitor Data Storage

The Host Monitor stores time series data and health data.

By default, the data is stored in /var/lib/cloudera-host-monitor/ on the Host Monitor host. You can change this by modifying the Host Monitor Storage Directory configuration. To change this configuration on an active system, follow the procedure in *Moving Monitoring Data on an Active Cluster*.

You can control how much disk space to reserve for Host Monitor data by changing the following configuration option:

- Time-series metrics and health data: Time Series Storage (firehose_time_series_storage_bytes - 10 GB default, 10 GB minimum)

For information about how metric data is stored in Cloudera Manager and how storage limits impact data retention, see *Data Granularity and Time-Series Metric Data*.

The default value is small, so you should examine disk usage after several days of activity to determine how much space they need. The Charts Library tab on the Cloudera Management Service page shows the current disk space consumed and its rate of growth, categorized by the type of data stored. For example, you can compare the space consumed by raw metric data to daily summaries of that data.

Viewing Host and Service Monitor Data Storage

The Cloudera Management Service page shows the current disk space consumed and its rate of growth, categorized by the type of data stored. For example, you can compare the space consumed by raw metric data to daily summaries of that data.

Procedure

1. Select ClustersCloudera Management Service.
2. Click the Charts Library tab.

Data Granularity and Time-Series Metric Data

The Service Monitor and Host Monitor store time-series metric data in a variety of ways.

When the data is received, it is written as-is to the metric store. Over time, the raw data is summarized to and stored at various data granularities. For example, after ten minutes, a summary point is written containing the average of the metric over the period as well as the minimum, the maximum, the standard deviation, and a variety of other statistics. This process is summarized to produce hourly, six-hourly, daily, and weekly summaries. This data summarization procedure applies only to metric data. When the Impala query and YARN application monitoring storage limit is reached, the oldest stored records are deleted.

The Service Monitor and Host Monitor internally manage the amount of overall storage space dedicated to each data granularity level. When the limit for a level is reached, the oldest data points at that level are deleted. Metric data for that time period remains available at the lower granularity levels. For example, when an hourly point for a particular time is deleted to free up space, a daily point still exists covering that hour. Because each of these data granularities consumes significantly less storage than the previous summary level, lower granularity levels can be retained for longer periods of time. With the recommended amount of storage, weekly points can often be retained indefinitely.

Some features, such as detailed display of health results, depend on the presence of raw data. Cluster utilization reports depend on hourly data being available for the selected time range. Charts built from weekly data might show a large gap between the last data point and the current time, as the next data point is not available yet. Other granularity levels may exhibit a similar gap, due in part to delays in the summarization process. These gaps may coincide with another one on the other side of the chart if metric history is too short to cover the selected time range. Health history is maintained by the event store dictated by its retention policies.

Moving Monitoring Data on an Active Cluster

You can change where monitoring data is stored on a cluster.

Basic: Changing the Configured Directory

1. Stop the Service Monitor or Host Monitor.
2. Save your old monitoring data and then copy the current directory to the new directory (optional).
3. Update the Storage Directory configuration option (firehose.storage.base.directory) on the corresponding role configuration page.
4. Start the Service Monitor or Host Monitor.

Advanced: High Performance

For the best performance, and especially for a large cluster, Host Monitor and Service Monitor storage directories should have their own dedicated spindles. In most cases, that provides sufficient performance, but you can divide your data further if needed. You cannot configure this directly with Cloudera Manager; instead, you must use symbolic links.

For example, if all your Service Monitor data is located in /data/1/service_monitor, and you want to separate your Impala data from your time series data, you could do the following:

1. Stop the Service Monitor.
2. Move the original Impala data in /data/1/service_monitor/impala to the new directory, for example /data/2/impala_data.
3. Create a symbolic link from /data/1/service_monitor/impala to /data/2/impala_data with the following command:

```
ln -s /data/2/impala_data /data/1/service_monitor/impala
```

4. Start the Service Monitor.

Host Monitor and Service Monitor Memory Configuration

You can configure Java heap size and non-Java memory size. The memory recommended for these configuration options depends on the number of hosts in the cluster, the services running on the cluster, and the number of monitored entities.

Monitored entities are the objects monitored by the Service Monitor or Host Monitor. As the number of hosts and services increases, the number of monitored entities also increases.

Aspects of your deployment, workload, and data storage that further increase the number of monitored entities include, without limitation: role instances, Kudu tablet replicas, Kafka producers, and monitored directories.

In addition to the memory configured, the Host Monitor and Service Monitor use the Linux page cache. Memory available for page caching on the Host Monitor and Service Monitor hosts improves performance.

Configuring Memory Allocations

To configure memory allocations, determine how many entities are being monitored and then consult the tables below for required and recommended memory configurations.

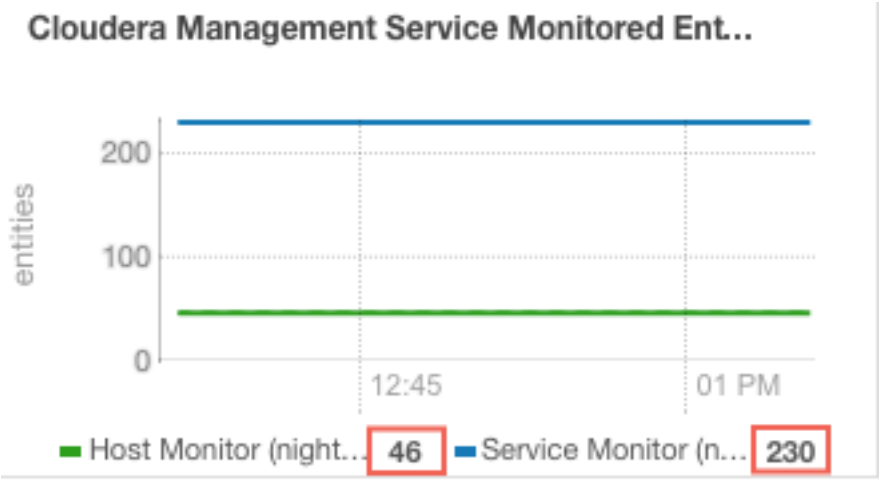
About this task

To determine the number of entities being monitored:

Procedure

1. Go to ClustersCloudera Management Service.
2. Locate the chart with the title Cloudera Management Service Monitored Entities.

   The number of monitored entities for the Host Monitor and Service Monitor displays at the bottom of the chart. In the following example, the Host Monitor has 46 monitored entities and the Service Monitor has 230 monitored entities.



3. Use the number of monitored entities for the Host Monitor to determine its memory requirements and recommendations in the tables below.

4. Use the number of monitored entities for the Service Monitor to determine its memory requirements and recommendations in the tables below.

   Clusters with HDFS, YARN, or Impala

   Use the recommendations in this table for clusters where the only services having worker roles are HDFS, YARN, or Impala.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-2,000	0-100	3 GB	18 GB
2,000-4,000	100-200	4.5 GB	18 GB
4,000-8,000	200-400	4.5 GB	36 GB
8,000-16,000	400-800	7.5 GB	36 GB
16,000-20,000	800-1,000	10.5 GB	36 GB

Clusters with HBase, Solr, Kafka, or Kudu

Use the recommendations when services such as HBase, Solr, Kafka, or Kudu are deployed in the cluster. These services typically have larger quantities of monitored entities.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-30,000	0-100	6 GB	36 GB
30,000-60,000	100-200	9 GB	36 GB
60,000-120,000	200-400	10.5 GB	36 GB
120,000-240,000	400-800	24 GB	60 GB

Additional tuning

Exact memory requirements on a given cluster are determined by numerous factors, and might change over time. Following the steps on this page again from time to time, or configuring limits higher than the recommendations herein, might become advisable (or necessary). You might notice different symptoms regarding the operation of Service Monitor or Host Monitor when this is the case:

- The Pause Duration health test shows Concerning or Bad status.
- Monitoring performance is unsatisfactory in any way.
- Logs frequently indicate pauses longer than 1000-2000 ms detected by JvmPauseMonitor, except when the message includes no GCs detected.



Important: Always verify that the above message is present in the logs before taking further action. If the JvmPauseMonitor lines indicate no GCs detected, or they are absent altogether, the root cause is likely unrelated to memory allocation.

If the recommended settings do not yield satisfactory results, Cloudera recommends using the values from the next row down until the problem is resolved. If this does not happen even with the highest recommended values, further increase is possible with no known limits other than the physical memory in the host. Set non-Java memory size to 100-150% of the Java heap size in this case.



Important: Using G1GC is strongly recommended for medium to large heap sizes.

Accessing Storage Using Microsoft ADLS Gen 2

These topics focused on Microsoft ADLS from the core Cloudera Enterprise documentation library can help you deploy, configure, manage, and secure clusters in the cloud.

They are listed by broad category:

Configuring Azure credentials for external table:

- To access data from external tables in the Azure filesystem, you need to configure Azure credentials by following these steps:

  1. Go to Cloudera Manager Administration External Accounts .
  2. Add credentials under **Azure Credentials**.

  This configuration is essential for seamless access to Azure-based external tables.

Note the following limitations:

- ADLS is not supported as the default filesystem. Do not set the default file system property (fs.defaultFS) to an abfss:// URI. You can use ADLS as secondary filesystem while HDFS remains the primary filesystem.
- Hadoop Kerberos authentication is supported, but it is separate from the Azure user used for ADLS authentication.
- Directory and file names should not end with a period. Paths that end in periods can cause inconsistent behavior, including the period disappearing. For more information, see HADOOP-15860.

Related Information

Inserting data into a Hive table created on Microsoft ADLS fails with an AzureADAuthenticator exception

Configuring OAuth in Data Hub

To connect a DataHub cluster to ADLS Gen2 with OAuth, you must configure the Hadoop CredentialProvider or core-site.xml directly. Although configuring core-site.xml is convenient, it is insecure since the contents of core-site.xml are not encrypted. For this reason, Cloudera recommends using a credential provider.

Before you start, ensure that you have configured OAuth for Azure.

Configuring OAuth with core-site.xml

Before you begin

Configuring your OAuth credentials in core-site.xml is insecure. Cloudera recommends that you only use this method for development environments or other environments where security is not a concern.

Perform the following steps to connect your cluster to ADLS Gen2:

Procedure

1. In the Cloudera Manager Admin Console, search for the following property: Cluster-wide Advanced Configuration Snippet (Safety Valve) for core-site.xml. .
2. Add the following properties and values:

Table 3: OAuth Properties

Name	Value
fs.azure.account.auth.type	OAuth

Name	Value
fs.azure.account.oauth.provider.type	org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider
fs.azure.account.oauth2.client.endpoint	Provide your tenant ID: https://login.microsoftonline.com/<*TENANT_ID*>/oauth2/token
fs.azure.account.oauth2.client.id	Provide your <*CLIENT_ID*>
fs.azure.account.oauth2.client.secret	Provide your <*CLIENT_SECRET*>

What to do next

In addition, you can also provide account-specific keys. For example, you might want to provide your storage account key to make use of replication.

Property	Value
fs.azure.account.key.<STORAGE_ACCOUNT>.dfs.core.windows.net	Provide your <Storage_Account_Key>

Configuring OAuth with the Hadoop CredentialProvider

Before you begin

A more secure way to store your OAuth credentials is with the Hadoop CredentialProvider. When you submit a job, reference the CredentialProvider, which then supplies the OAuth information. Unlike the core-site.xml, the credentials are not stored in plain text.

The following steps describe how to create a credential provider and how to reference it when submitting jobs:

Procedure

1. Create a password for the Hadoop Credential Provider and export it to the environment:

```
export HADOOP_CREDSTORE_PASSWORD=PASSWORD
```

2. Provision the credentials by running the following commands:

```
hadoop credential create dfs.adls.oauth2.client.id -provider jceks://hdfs/
user/USER_NAME/adls2keyfile.jceks -value CLIENT ID
hadoop credential create dfs.adls.oauth2.credential -provider jceks://h
dfs/user/USER_NAME/adls2keyfile.jceks -value CLIENT SECRET
hadoop credential create dfs.adls.oauth2.refresh.url -provider jceks://
hdfs/user/USER_NAME/adls2keyfile.jceks -value REFRESH URL
```

   You can omit the -value option and its value and the command will prompt the user to enter the value.

   For more details on the hadoop credential command, see Credential Management (Apache Software Foundation).

3. Export the password to the environment:

```
export HADOOP_CREDSTORE_PASSWORD=password
```

4. Reference the credential provider on the command line when you submit a job:

```
hadoop <COMMAND>
     -Ddfs.adls.oauth2.access.token.provider.type=ClientCredential \
     -Dhadoop.security.credential.provider.path=jceks://hdfs/user/USER_NAM
E/adls-cred.jceks \
     abfs[s]://<file_system>@<account_name>.dfs.core.windows.net/<path>/
<file_name>
```

Configuring Built-in TLS Acceleration

For ADLS Gen2, TLS is enabled by default using the Java implementation of TLS. For better performance, you can use the built-in OpenSSL implementation of TLS.

Perform the following steps to use the built-in OpenSSL implementation of TLS:

1. Verify the location of the OpenSSL libraries on the hosts with the following command:

   ```
   whereis libssl
   ```

2. In the Cloudera Manager Admin Console, search for the following property: Gateway Client Environment Advanced Configuration Snippet (Safety Valve) for hadoop-env.sh.
3. Add the following parameter to the property:

   ```
   HADOOP_OPTS="-Dorg.wildfly.openssl.path=<path to OpenSSL libraries> ${HA
   DOOP_OPTS}"
   ```

   For example, if the OpenSSL libraries are in /usr/lib64, add the following parameter:

   ```
   HADOOP_OPTS="-Dorg.wildfly.openssl.path=/usr/lib64 ${HADOOP_OPTS}"
   ```

4. Save the change.
5. Search for the following property: HDFS Client Environment Advanced Configuration Snippet (Safety Valve) for hadoop-env.sh
6. Add the following parameter to the property:

   ```
   HADOOP_OPTS="-Dorg.wildfly.openssl.path=<PATH TO OPENSSL LIBRARIES> ${HADO
   OP_OPTS}"
   ```

   For example, if the OpenSSL libraries are in /usr/lib64, add the following parameter:

   ```
   HADOOP_OPTS="-Dorg.wildfly.openssl.path=/usr/lib64 ${HADOOP_OPTS}"
   ```

7. Save the change.
8. Restart the stale services.
9. Deploy the client configurations.
10. Verify that you configured built-in TLS acceleration successfully by running the following command from any host in the cluster:

    ```
    hadoop fs -ls abfss://<container>@<account>.dfs.core.windows.net/
    ```

    A message similar to the following should appear:

    ```
    org.wildfly.openssl.SSL init
    INFO: WFOPENSSL0002 OpenSSL Version OpenSSL 1.0.1e-fips 11 Feb 2013
    ```

    The message may differ slightly depending on your operating system and OpenSSL version.

Importing Data into Microsoft Azure Data Lake Store (Gen1 and Gen2) Using Sqoop

Microsoft Azure Data Lake Store (ADLS) is a cloud object store designed for use as a hyper-scale repository for big data analytic workloads. ADLS acts as a persistent storage layer for CDH clusters running on Azure.

There are two generations of ADLS, Gen1 and Gen2. You can use Apache Sqoop with both generations of ADLS to efficiently transfer bulk data between these file systems and structured datastores such as relational databases. For more information on ADLS Gen 1 and Gen 2, see:

- Microsoft ADLS Gen1 documentation
- Microsoft ADLS Gen2 documentation

You can use Sqoop to import data from any relational database that has a JDBC adaptor such as SQL Server, MySQL, and others, to the ADLS file system.



Note: Sqoop export from the Azure files systems is not supported.

The major benefits of using Sqoop to move data are:

- It leverages RDBMS metadata to get the column data types
- It ensures fault-tolerant and type-safe data handling
- It enables parallel and efficient data movement

Prerequisites

The configuration procedure presumes that you have already set up an Azure account, and have configured an ADLS Gen1 store or ADLS Gen2 storage account and container. See the following resources for information:

- Microsoft ADLS Gen1 documentation
- Microsoft ADLS Gen2 documentation
- Hadoop Azure Data Lake Support

Authentication

To connect a cluster to ADLS with OAuth, you must configure the Hadoop CredentialProvider or core-site.xml directly. Although configuring the core-site.xml is convenient, it is insecure, because the contents of core-site.xml configuration file are not encrypted. For this reason, Cloudera recommends using a credential provider. For more information, see Configuring OAuth in CDH .

You can also pass the credentials by providing them on the Sqoop command line as part of the import command.

```
sqoop import
-Dfs.azure.account.auth.type=...
-Dfs.azure.account.oauth.provider.type=...
-Dfs.azure.account.oauth2.client.endpoint=...
-Dfs.azure.account.oauth2.client.id=...
-Dfs.azure.account.oauth2.client.secret=...
```

For example:

```
sqoop import
-Dfs.azure.account.oauth2.client.endpoint=https://login.microsoftonline
.com/$TENANT_ID/oauth2/token
-Dfs.azure.account.oauth2.client.id=$CLIENT_ID
-Dfs.azure.account.oauth2.client.secret=$CLIENT_SECRET
-Dfs.azure.account.auth.type=OAuth
-Dfs.azure.account.oauth.provider.type=org.apache.hadoop.fs.azurebfs.oauth2.
ClientCredsTokenProvider
```

Sqoop Import into ADLS

To import data into ADLS from diverse data sources, such as a relational database, enter the Sqoop import command on the command line of your cluster. Make sure that you specify the Sqoop connection to the data source you want to import.

If you want to enter a password for the data source, use the -P option in the connection string. If you want to specify a file where the password is stored, use the --password-file option.

```
sqoop import
-Dfs.azure.account.auth.type=...
-Dfs.azure.account.oauth.provider.type=...
-Dfs.azure.account.oauth2.client.endpoint=...
-Dfs.azure.account.oauth2.client.id=...
-Dfs.azure.account.oauth2.client.secret=...
--connect... --username... --password... --table... --target-dir... --split-
by...
```

ABFS example:

```
sqoop import
-Dfs.azure.account.oauth2.client.endpoint=https://login.microsoftonline
.com/$TENANT_ID/oauth2/token
-Dfs.azure.account.oauth2.client.id=$CLIENT_ID
-Dfs.azure.account.oauth2.client.secret=$CLIENT_SECRET
-Dfs.azure.account.auth.type=OAuth
-Dfs.azure.account.oauth.provider.type=org.apache.hadoop.fs.azurebfs.oauth2.
ClientCredsTokenProvider
--connect $CONN --username $USER --password $PWD --table $TABLENAME --targe
t-dir abfs://$CONTAINER$ACCOUNT.dfs.core.windows.net/$TARGET-DIRECTORY --s
plit-by $COLUMN_NAME
```

ADLS example:

```
sqoop import
-Dfs.adl.oauth2.refresh.url=https://login.windows.net/$TENANT_ID/oauth2/toke
n
-Dfs.adl.oauth2.client.id=$CLIENT_ID
-Dfs.adl.oauth2.credential=$CLIENT_SECRET
-Dfs.adl.oauth2.access.token.provider.type=ClientCredential
--connect $CONN --username $USER --password $PWD --table $TABLENAME --targe
t-dir adl://$TARGET-ADDRESS/$TARGET-DIRECTORY --split-by $COLUMN_NAME
```

Configuration Change Audit API

Learn about the Configuration Change Audit Application Programming Interface (API) in Cloudera Manager, including its operational capabilities, query parameters, authorization model, and behavioral semantics. Understand the business purpose and high-level architecture of the Configuration Change Audit REST API in Cloudera Manager.

Overview

Cloudera Manager maintains configuration change history through its existing revision and audit infrastructure. Historically, administrators accessed this history through the Config History user interface (UI). External security, governance, and compliance systems lacked a supported Representational State Transfer (REST) API to query configuration differences programmatically.

The Configuration Change Audit API introduces a dedicated REST endpoint:

```
GET /api/v58/audits/configChanges
```

The API exposes configuration change records in a structured, filterable format that includes the following details:

- The user who made the change
- The timestamp when the change occurred

- The revision identifier (revisionId)
- The affected cluster, service, or role
- The configuration parameter name
- The previous configuration value
- The new configuration value

Key Benefits

Understand the evolution of configuration change tracking in Cloudera Manager and how the dedicated Configuration Audit API simplifies integration with enterprise security and automation platforms.

Legacy Configuration racking

Cloudera Manager recorded configuration changes using Hibernate Envers and dedicated revision tables. While the Cloudera Manager Admin Console consumed this information in the Config History UI, external consumers had to assemble configuration audit details indirectly. External tools had to combine raw event streams with low-level audit records, which created maintenance overhead and led to inconsistent parsing.

How the API Resolves Integration Challenges

The new API provides a supported, programmatic interface that allows external security and automation tools to extract configuration audit records directly. The API addresses several core operational use cases, including the following scenarios:

- Compliance audits: Query all configuration changes made across managed clusters within the past 30 days.
- Security investigations: Determine which user changed sensitive configuration parameters, such as dfs.replication.
- Security Information and Event Management (SIEM) integration: Poll configuration changes regularly and correlate configuration modifications with external security events.
- Incident investigation: Inspect previous configuration values before an outage or regression occurred.

Key Features and Functional Capabilities

Understand the core design objectives, query parameters, and security features supported by the Configuration Change Audit API.

The Configuration Change Audit API provides the following core capabilities:

- Stable REST endpoint: Access configuration audit events programmatically through GET /api/v58/audits/configChanges.
- Hierarchical data model: Organizes change records logically across revisions, clusters, services, and modified entities.
- Timestamp-based filtering: Enables precise historical windowing using standardized ISO timestamps.
- Granular resource scoping: Restricts queries to target clusters, specific services, or individual configuration parameters.
- FIQL query syntax: Leverages Feed Item Query Language (FIQL) expressions to execute complex, multi-attribute searches.
- Revision-level pagination: Groups response pages by complete revision boundaries to preserve transaction integrity.
- Role-based access control (RBAC): Restricts access to audit history using dedicated Cloudera Manager user authorities.
- Sensitive data redaction: Automatically masks passwords and secret parameters in parity with Config History UI redaction policies.
- Direct revision lookup: Fetches specific change events directly using unique revisionId identifiers.

API Query Parameters

Review the supported query parameters, their default values, and available filtering options.

Every query parameter for GET /api/v58/audits/configChanges is optional and includes a default value.

Parameter	Default Value
revisionId	None
clusterName	None
serviceName	None
configParameter	None
startTime	1 day before endTime
endTime	Current time (now)
limit	100
beforeRevisionId	None
query	None

Paginating Audit Results

Understand cursor-based navigation using beforeRevisionId to retrieve large audit trails safely without splitting change sets.

The Configuration Change Audit API uses revision-level pagination rather than record-level offsets. When an administrator saves a configuration change in Cloudera Manager, that single transaction can update multiple parameters under one revision ID. Paginating strictly across revision boundaries ensures that Cloudera Manager never fragments a single logical save operation across separate API response pages.

```
Request 1: GET .../configChanges?limit=100
           (Returns revisions 500 down to 401. Smallest revisionId = 401)
     #
Request 2: GET .../configChanges?limit=100&beforeRevisionId=401
           (Returns revisions 400 down to 301. Smallest revisionId = 301)
     #
Request 3: GET .../configChanges?limit=100&beforeRevisionId=301
           (Returns revisions 300 down to 201)
```

How to Paginate Through Results

To paginate through audit results:

1. Send an initial GET request without the beforeRevisionId parameter.
2. Identify the lowest revisionId value in the returned response list.
3. Pass that value as the beforeRevisionId query parameter in your subsequent request.
4. Repeat this process until the API returns an empty list.

Security Governance and Redaction

Review the authentication privileges, role requirements, and automated credential redaction policies enforced by the endpoint.

Role-Based Access Control (RBAC)

Cloudera Manager controls access to this endpoint through the dedicated AUTH_AUDITS permission authority.

Cloudera Manager User Role	Has AUTH_AUDITS Authority
ROLE_ADMIN	Yes

Cloudera Manager User Role	Has AUTH_AUDITS Authority
ROLE_AUDITOR	Yes
ROLE_USER and lower roles	No

If a user without the AUTH_AUDITS authority submits a request to the endpoint, Cloudera Manager rejects the request with an HTTP 403 Forbidden response.

Sensitive Data Redaction

To prevent credential leaks, the API enforces the same redaction rules as the Config History UI. Sensitive values such as passwords, private keys, Kerberos keytab contents, and confidential credentials, automatically appear as masked strings in both the oldValue and newValue fields.

Configuration Audit Scope

Identify the configuration objects included in the audit stream and the administrative settings excluded from this endpoint.

Supported Configuration Objects

The API audits configuration changes associated with the following resources:

- Service-level configurations: Global service properties, such as default HDFS replication settings and YARN cluster memory allocations.
- Role-level configurations: Process-specific parameters, such as NameNode Java Virtual Machine (JVM) heap limits.
- Role Config Groups (RCG): Specialized configuration group property overrides.
- Host-level configurations: Host-specific Cloudera Manager Agent parameters.

Excluded Configuration Objects

The following configuration resources fall outside the scope of this endpoint:

- Cloudera Manager server settings: Internal Cloudera Manager administrative configurations use a separate auditing path.
- External account configurations: External authentication and third-party credential associations fall outside cluster configuration auditing.

API Behavior and Validation Rules

Learn how the API evaluates query parameters, handles boundary conditions, parses parameter changes, and validates input requests.

Parameter Precedence (revisionId)

When you specify revisionId, Cloudera Manager treats it as the highest-priority filter. The API ignores all additional filtering parameters in the request (such as clusterName, serviceName, star tTime, endTime, limit, beforeRevisionId, or query) without returning an error.

Handling Empty Query Parameters

The API treats empty parameter keys or empty assigned values as omitted parameters. The following requests produce identical results:

- GET /api/v58/audits/configChanges?clusterName
- GET /api/v58/audits/configChanges?clusterName=
- GET /api/v58/audits/configChanges



Important: Cloudera Manager does not treat an empty parameter as an invalid filter, an empty string query, or a zero-result constraint. The API applies standard defaults for time range and pagination.

Half-Open Time Range Evaluation

The API calculates time boundaries using a standard half-open interval:

```
[startTime,endTime)
```

Cloudera Manager includes revisions where:

```
timestamp # startTimeANDtimestamp < endTime
```

Deletion and Reset Semantics

When an administrator removes an explicit role-level or service-level configuration override, the audit record returns an oldValue without a corresponding newValue:

```
{
  "configParameter": "dfs.datanode.data.dir",
  "oldValue": "/data/dfs/dn"
}
```

This payload indicates that the administrator deleted an override at that specific configuration level, causing the parameter to inherit its default value from the parent level.

Addition and Override Semantics

When an administrator sets an explicit value for a parameter that previously had no custom value, the audit record returns a newValue without an oldValue:

```
{
  "configParameter": "dfs.replication",
  "newValue": "3"
}
```

This payload indicates that the administrator added a new, explicit configuration value at the audited hierarchy level.

Duplicate Parameter Validation

The API allows each supported query parameter to appear only once per request. Supplying the same query parameter multiple times causes Cloudera Manager to reject the call with an HTTP 400 Bad Request status:

```
GET /api/v58/audits/configChanges?clusterName=ClusterA&clusterNa
me=ClusterB
HTTP/1.1 400 Bad Request
```