cloudera
Ask Bigger Questions

Cloudera Manager Administration Guide

Important Notice

(c) 2010-2013 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, Cloudera Impala, and any other product or service names or slogans contained in this document are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder.

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners. Reference to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

Cloudera, Inc.
1001 Page Mill Road, Building 2
Palo Alto, CA 94304-1008
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-362-0488
www.cloudera.com

Release Information

Version: 4.7.3

Date: October 31, 2013

Table of Contents

About this Guide

This guide is for system administrators who need to manage a Cloudera Manager server installation. This guide covers adding and managing Cloudera Manager users, adding or upgrading licenses, configuring TLS security, configuring the Alert Publisher, and other similar features.

Managing the Cloudera Manager Server and Agents

This section covers information on managing the Cloudera Manager server and the agents that run on each node of the cluster. This section covers the following topics:

- Starting and Logging into the Admin Console
- Stopping or Restarting the Cloudera Manager Server on page 9
- Stopping or Restarting Cloudera Manager Agents on page 9
- Configuring Agent Heartbeat and Health Status Options on page 10
- Configuring the Ports for the Admin Console and Agents on page 11
- Viewing the Cloudera Manager Server and Agent Logs on page 11

Stopping or Restarting the Cloudera Manager Server

You can stop the Cloudera Manager server (for example, to perform maintenance on its host) 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 without affecting other services:

```
service cloudera-scm-server stop
```

To restart it:

```
service cloudera-scm-server start
```

Note:

If you are intending to perform an upgrade of Cloudera Manager, then you should stop the management service (through the Admin Console) prior to stopping the server.

Stopping or Restarting Cloudera Manager Agents

Usually (during an upgrade to a new version of Cloudera Manager, for example) you want to stop or restart the Agents while leaving the processes they manage running. To do this, use one of the following commands on every Agent host.

- To stop the Agent itself, but leave the processes it manages running:

```
$ sudo service cloudera-scm-agent stop
```

- To restart a running Agent without restarting any of the processes it manages:

```
$ sudo service cloudera-scm-agent restart
```

If you want to stop or restart the Agents themselves and the services they manage, use one of the following commands on every Agent host.

- To stop the Agent and the processes it manages:

```
$ sudo service cloudera-scm-agent hard_stop
```

- To restart the running Agent and the processes it manages:

```
$ sudo service cloudera-scm-agent hard_restart
```

When an Agent is stopped using either of the `stop` or `hard_stop` commands, you cannot use either of the `restart` or `hard_restart` commands to start it. You must use the following `start` command to start a stopped agent regardless of how you stopped it:

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

Configuring Agent Heartbeat and Health Status Options

You can configure the Cloudera Manager Agent heartbeat interval and timeouts to trigger changes in Agent health status.

To configure Agent heartbeat and health status options:

1. From the **Administration** tab, select **Settings**.
2. Under the **Performance** category, set the following option:

Setting	Description
Send Agent Heartbeat Every ___ second(s)	The interval between each heartbeat that is sent from Cloudera Manager Agents to the Cloudera Manager Server.

3. Under the **Threshold** category, set the following options:

Setting	Description
Set health status to Concerning if the Agent heartbeats fail ____ time(s)	If an Agent fails to send this number of expected consecutive heartbeats to the Server, a **Concerning** health status is assigned to that Agent.
Set health status to Bad if the Agent heartbeats fail ____ time(s)	If an Agent fails to send this number of expected consecutive heartbeats to the Server, a **Bad** health status is assigned to that Agent.

4. Click **Save Changes**.

For information about health status, see Viewing Service Status.

Configuring the Ports for the Admin Console and Agents

You can configure the HTTP and HTTPS ports you want to use for the Cloudera Manager Admin Console and Agents.

To configure the ports for the Cloudera Manager Admin Console and Agents:

1. From the **Administration** tab, select **Settings**.
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 via the Admin Console.
HTTPS Port for Admin Console	Specify the HTTPS port to use to access the Server via 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 by typing the following command on the Cloudera Manager Server host:

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

Viewing the Cloudera Manager Server and Agent Logs

To help you troubleshoot problems, you can view the Cloudera Manager Server log.

To view the Cloudera Manager Server log:

1. Pull down the **Diagnose** menu, and select **Server Log**.

To view the Cloudera Manager Agent log:

1. Click the **Hosts** tab.
2. Click the link for the host where you want to see the Agent logs.
3. In the **Details** panel, click the **Details** link in the **Host Agent** column.
4. Click the **Agent Log** link.

Note:

You can also view the Cloudera Manager Server log at `/var/log/cloudera-scm-server/cloudera-scm-server.log` on the Server host or the Cloudera Manager Agent log at `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts for information about the problems.

Cloudera Manager Users and Authentication

This chapter covers managing user accounts, and configuring external authentication.

Cloudera Manager User Accounts

Cloudera Manager user accounts allow users to log into the Cloudera Manager Admin Console. User authentication can be done through a local database, through an external LDAP directory server (Active Directory or OpenLDAP-compatible), or through an external authentication program of your own choosing.

Cloudera Manager users are managed through the **Administration** **Users** page (accessed from the **Administration** tab).

User accounts added from an LDAP directory or other external authentication mechanism will have **External** in the **User Type** column shown under the Users tab. Users in the local database will have **Cloudera Manager** as the user type. See Configuring External Authentication for information on configuring Cloudera Manager to use an external LDAP directory or other authentication program for user authentication.

User accounts can optionally have Administrator privileges:

- Administrator privileges: Allows the user to add, change, delete, and configure services or administer user accounts. Also, even if you are using an external authentication mechanism for user authentication, users with Administrator privileges will also be able to log in to Cloudera Manager using their local Cloudera Manager username and password. (This prevents the system from locking everyone out if the external authentication settings get misconfigured.)
- No Administrator privileges: User accounts that don't have Administrator privileges can view services and monitoring information but they cannot add services or take any actions that affect the state of the cluster.

When you are logged in to the Cloudera Manager Admin Console, the user name you are logged in as is shown on the top navigation bar — for example, if you are logged in as *admin* you will see this:

.

Changing Your Password

> **Important:**
>
> As soon as possible after running the installation wizard and beginning to use Cloudera Manager, you should use the following procedure to change the password for the default `admin` account, if you have not already done so.

To change the password for the logged-in account:

1. Logged in as `admin` pull down user menu and select the **Change Password** option.
2. Enter a new password twice and then click **Submit**.

Adding Cloudera Manager User Accounts

To add a Cloudera Manager user account:

1. From the **Administration** tab, select **Users** .
2. Click the **Add User** button.
3. Enter a username and password.
4. To grant Administrator privileges to the user account, select **Add Administrator Privileges**.
5. Click **Submit**.

Users accounts created in this way will show **Cloudera Manager** in the User Type column.

Changing the Privileges and Password for an Account

To change the privileges for an account:

1. Click the checkbox to the left to select the user account.
2. Click the **Add Administrator Privileges** or **Remove Administrator Privileges** button.

To change an account password:

1. Click the **Change Password** button.
2. Type the new password and repeat it to confirm.
3. Click the **Submit** button to make the change.

Deleting an Account

To delete an account:

1. Select the user account.
2. Click the **Delete** button. (Note that there is no confirmation of the action.)

Configuring External Authentication

> **Note: This feature is available only with Cloudera Enterprise.**
>
> The feature described in this section is not available in Cloudera Manager with Cloudera Standard.
>
> If you have been using the Cloudera Enterprise Trial Edition, this feature will no longer be available after your trial license expires.
>
> To obtain a license for Cloudera Enterprise, please contact sales@cloudera.com. When you install your Enterprise license, this feature will be enabled.

Cloudera Manager provides several different mechanisms for authenticating users for Cloudera Manager. You can enter users into Cloudera Manager's own database (the default) or configure Cloudera Manager to authenticate against an external authentication service. This can be an LDAP server (Active Directory or an OpenLDAP compatible directory) or you can specify another external service.

Further, you can configure Cloudera Manager so that it can use both methods of authentication (internal database vs. external service), and you can determine the order in which it performs these searches. You can also restrict login access to members of specific groups, and can specify groups whose members will automatically be given administrator access to Cloudera Manager.

For an OpenLDAP compatible directory, you have several options for searching for users and groups:

- You can specify a single base Distinguished Name (DN) and then provide a "Distinguished Name Pattern" to use to match a specific user in the LDAP directory.
- Search filter options let you search for a particular user based on somewhat broader search criteria – for example Cloudera Manager users could be members of different groups or organizational units (OUs), so a

single pattern won't find all those users. Search filter options also let you find all the groups to which a user belongs, to help determine if that user should have login or admin access.

To configure an external authentication service for Cloudera Manager user authentication:

1. From the **Administration** tab, select the **Settings** .
2. In the left-hand column, select the **External Authentication** category.
3. Select the order in which Cloudera Manager should attempt its authentication (**Authentication Backend Order**). Here you can choose to authenticate users using just one of the methods (using Cloudera Manager's own Database is the default), or you can set it so that if the user cannot be authenticated by the first method, it will attempt using the second method. Note that if you select **External Only**, users who are administrators in the Cloudera Manager database will still be able to log in with their database password. This is to prevent the system from locking everyone out if the authentication settings get misconfigured — such as with a bad LDAP URL.
4. Go to the section below for the type of authentication you want to configure, and follow the steps to set the properties appropriately.

Configure User Authentication Using Active Directory

1. For **External Authentication Type** select Active Directory.
2. Provide the URL of the Active Directory server.
3. Provide the NT domain to authenticate against.
4. Optionally, provide a comma-separated list of LDAP group names in the **LDAP User Groups** property. If this list is provided, only users who are members of one or more of the groups in the list will be allowed to log into Cloudera Manager. If this property is left empty, *all* authenticated LDAP users will be able to log into Cloudera Manager. For example, if there is a group called "CN=ClouderaManagerUsers,OU=Groups,DC=corp,DC=com", add the group name `ClouderaManagerUsers` to the **LDAP User Groups** list to allow members of that group to log in to Cloudera Manager. The group names are case-sensitive.
5. In the **LDAP Administrator Groups** property you can provide a list of groups whose members should be given administrator access when they log in to Cloudera Manager. (Note that admin users must also be a member of at least one of the groups specified in the **LDAP User Groups** property or they will not be allowed to log in.) If this is left empty, then no users will be granted administrator access automatically at login — administrator access will need to be granted manually by another administrator.

Configure User Authentication Using an OpenLDAP-compatible Server

1. For **External Authentication Type** select **LDAP**.
2. Provide the URL of the LDAP server and (optionally) the base Distinguished Name (DN) (the search base) as part of the URL — for example `ldap://ldap-server.corp.com/dc=corp,dc=com`.
3. **If your server does NOT allow anonymous binding:** Provide the user DN and password to be used to bind to the directory. These are the **LDAP Bind User Distinguished Name** and **LDAP Bind Password** properties. By default, Cloudera Manager assumes anonymous binding.
4. To use a single "Distinguished Name Pattern," provide a pattern in the **LDAP Distinguished Name Pattern** property.

   Use {`0`} in the pattern to indicate where the username should go. For example, to search for a distinguished name where the the uid attribute is the username, you might provide a pattern similar to `uid={0},ou=People,dc=corp,dc=com`. Cloudera Manager substitutes the name provided at login into this pattern and performs a search for that specific user. So if a user provides the username "foo" at the Cloudera Manager login page, Cloudera Manager will search for the DN `uid=foo,ou=People,dc=corp,dc=com`.

Note that if you provided a base DN along with the URL, the pattern only needs to specify the rest of the DN pattern. For example, if the URL you provide is `ldap://ldap-server.corp.com/dc=corp,dc=com`, and the pattern is `uid={0},ou=People`, then the search DN will be `uid=foo,ou=People,dc=corp,dc=com`.

5. You can also search using User and/or Group search filters, using the **LDAP User Search Base**, **LDAP User Search Filter**, **LDAP Group Search Base** and **LDAP Group Search Filter** settings. These allow you to combine a base DN with a search filter to allow a greater range of search targets.

   For example, if you want to authenticate users who may be in one of multiple OUs, the search filter mechanism will allow this. You can specify the User Search Base DN as `dc=corp,dc=com` and the user search filter as `uid={0}`. Then Cloudera Manager will search for the user anywhere in the tree starting from the Base DN. Suppose you have two OUs — ou=Engineering and ou=Operations — Cloudera Manager will find User "foo" if it exists in either of these OUs, i.e. `uid=foo,ou=Engineering,dc=corp,dc=com` or `uid=foo,ou=Operations,dc=corp,dc=com`.

   You can use a user search filter along with a DN pattern, so that the search filter provides a fallback if the DN pattern search fails.

   The Groups filters let you search to determine if a DN or user name is a member of a target group. In this case, the filter you provide can be something like `member={0}` where {0} will be replaced with the **DN** of the user you are authenticating. For a filter requiring the user name, {1} may be used, as `memberUid={1}`. This will return a list of groups this user belongs to, which will be compared to the list in the **LDAP User Groups** and **LDAP Administrator Groups** properties (discussed previously in the section about Active Directory).

Configure Cloudera Manager to use LDAPS instead of LDAP:

If the LDAP server's certificate has been signed by a trusted Certificate Authority (ie., VerSign, GeoTrust, etc.) the following steps may not be necessary.

1. Copy the CA certificate file (`ca.cer`, etc.) to the Cloudera Manager server.
2. Import the CA certificate(s) from the CA certificate file to the local keystore.

   Example: `/usr/java/latest/bin/keytool -import -alias <nt_domain_name> -keystore /usr/java/latest/jre/lib/security/cacerts -file <path_to_cert>`

   > **Note:**
   >
   > - The default password for the cacerts store is **changeit**.
   > - The alias can be any name (not just the domain name).

3. Configure the LDAP URL in the Cloudera Manager configuration to use `ldaps://<ldap_server>` instead of `ldap://<ldap_server>`

Configure User Authentication Using an External Program

You can configure Cloudera Manager to use an external authentication program of your own choosing. Typically, this may be a custom script that interacts with a custom authentication service. Cloudera Manager will call the external program with the user name as the first command line argument. The password is passed over `stdin`. Cloudera Manager assumes the program will return the following exit codes:

- 0 for the successful authentication of a regular user
- 1 for the successful authentication of an admin user
- a negative value for failure to authenticate.

1. For **External Authentication Type** select **External Program**.
2. Provide a path to the external program in the **External Authentication Program Path** property.

Configuring User Authentication Using SAML

Cloudera Manager supports a number of authentication mechanisms, both internal and using external services. This includes the Security Assertion Markup Language (SAML), an XML-based open standard data format for exchanging authentication and authorization data between parties, in particular, between an identity provider (IDP) and a service provider (SP).

The SAML specification defines three roles: the principal (typically a user), the identity provider, and the service provider. In the use case addressed by SAML, the principal (user agent) requests a service from the service provider. The service provider requests and obtains an identity assertion from the identity provider. On the basis of this assertion, the service provider can make an access control decision - in other words it can decide whether to perform some service for the connected principal.

The primary SAML use case is called Web Browser Single Sign-On (SSO). A user wielding a user agent (usually a web browser) requests a web resource protected by a SAML service provider. The service provider, wishing to know the identity of the requesting user, issues an authentication request to a SAML identity provider through the user agent.

In the context of this terminology, Cloudera Manager operates as a Service Provider.

Overview and Assumptions

This document assumes that you are familiar with SAML and SAML configuration in a general sense, and that you have a functioning IDP already deployed. It will only discuss the Cloudera Manager specific part of the configuration process.

At a high level, setting up Cloudera Manager to use SAML is a three stage process.

- Configure Cloudera Manager to act as a Service Provider.
- Configure your IDP to recognize Cloudera Manager as a valid SP.
- Confirm that Cloudera Manager can correctly authenticate users with the IDP.

Prerequisites and Preparation

You will need to prepare the following files and information, and provide these to Cloudera Manager:

- A Java Keystore containing:
  - A private key for CM to use to sign/encrypt SAML messages
  - Any public certificates needed to verify the sign/encrypt key used by your Identity Provider
- The SAML metadata XML file from your IDP
- The entity ID that should be used to identify the Cloudera Manager instance
- How the user ID is passed in the SAML authentication response:
  - As the NameID
  - As an attribute. If so, what identifier is used.
- The method by which the Cloudera Manager role will be established (admin vs. regular user):
  - From an attribute in the authentication response:
    - What identifier will be used for the attribute
    - What values will be passed to indicate each role
  - From an external script that will be called for each use:

- The script takes user ID as $1
- The script sets an exit code to reflect assigned role
  - 0 = admin
  - 1 = regular user
  - -1 = failure

Configuring Cloudera Manager

1. Start the server normally and log in using an Admin account.
2. From the Administration tab, select Settings, then External Authentication.
3. Set the External Authentication Type to SAML (note that the Authentication Backend Order is ignored for SAML).
4. Set the metadata XML file path to point to the IDP's metadata file.
5. Set the keystore file path to point to the Java Keystore prepared earlier.
6. Set the Keystore's password.
7. Set the alias used to identify the private key for CM to use.
8. Set the private key's password.
9. Set the entity ID if necessary:
   - If there is more than one CM instance being used with the same IDP (each instance needs a different entity ID).
   - If entity IDs are assigned by organizational policy.
10. Set whether the user ID will be obtained from the NameID or an attribute.
11. If an attribute will be used, set the attribute name if necessary. The default value is the normal OID used for user IDs and so may not need to be changed.
12. Set whether the role assignment will be done from an Attribute or an external script:
    - If an attribute will be used, set the attribute name if necessary. The default value is the normal OID used for OrganizationalUnits and so may not need to be changed.
    - If an attribute will be used, set what attribute values will be used to indicate admins vs regular users.
    - If an external script will be used, set the path to that external script. Make sure that the script is executable (Note that an executable binary is fine - it doesn't need to be a literal shell script).
13. Save the changes:
    - CM will run a set of validations that ensure it can find the metadata XML and the keystore, and that the passwords are correct. If you see a validation error, please correct the problem before proceeding.
14. Restart the Cloudera Manager server

Configuring the IDP

After Cloudera Manager is restarted, it will attempt to redirect to the IDP's login page instead of showing the normal CM page. This may or may not succeed, depending on how the IDP is configured. In either case, the IDP will need to be configured to recognize CM before authentication will actually succeed. The details of this process are specific to each IDP implementation and cannot be described here - refer to your IDP's documentation for details

1. Download Cloudera Manager's SAML metadata xml file:

  - `http://<hostname>:7180/saml/metadata`

2. Inspect the metadata file and ensure that any URLs contained in the file can be resolved by users' web browsers. The IDP will redirect web browsers to these URLs at various points in the process. If the browser cannot resolve them, authentication will fail.
   - If the URLs are incorrect, you can manually fix the xml file or set the Entity Base URL in the CM configuration to the right value, and then re-download the file.
3. Provide this metadata file to your IDP using whatever mechanism your IDP provides.
4. Ensure that the IDP has access to whatever public certificates are necessary to validate the private key that was provided to Cloudera Manager earlier.
5. Ensure that the IDP is configured to provide the User ID and Role using the attribute names that Cloudera Manager was configured to expect, if relevant.
6. Ensure the changes to the IDP configuration have taken effect (a restart may be necessary).

Verifying that authentication and authorization can be completed

1. Return to Cloudera Manager and refresh the login page.
2. Attempt to log in with credentials for a user that is entitled as an admin or a regular user.
3. The authentication should complete and you should see the Cloudera Manager home page.
4. If authentication fails, you will see an IDP provided error message. Cloudera Manager is not involved in this part of the process, and you must ensure the IDP is working correctly to complete the authentication.
5. If authentication succeeds but the user is not authorized to use Cloudera Manager (as an Admin or a regular User), they will be taken to a special error page by Cloudera Manager that explains the situation.
   - If an user who should be authorized sees this error, then you will need to verify their role configuration, and ensure that it is being properly communicated to CM, whether by attribute or external script. The CM log will provide details on failures to establish a user's role. If any errors occur during role mapping, CM will assume the user is unauthorized.

Interoperability Notes

- Cloudera Manager supports both SP-initiated and IDP-initiated Single-Sign-On.
- The Logout action in Cloudera Manager will send a Single-Logout request to the IDP.
- SAML authentication has been tested with specific configurations of Siteminder and Shibboleth. While SAML is a standard, there is a great deal of variability in configuration between different IDP products, so it is possible that other IDP implementations, or other configurations of Siteminder and Shibboleth, may not interoperate correctly with Cloudera Manager.

Configuring TLS Security for Cloudera Manager

- **Important:** If you want to add new hosts after performing the following procedures to enable TLS, you must disable TLS and then configure TLS for each new host. For more information, see Adding a Host to the Cluster.

- **Important:** Cloudera strongly recommends that you set up a fully-functional CDH cluster and Cloudera Manager before you begin configuring it to use TLS.

Transport Layer Security (TLS) provides encryption and authentication in the communications between the Cloudera Manager Server and Agents. Encryption prevents snooping of communications, and authentication helps prevent malicious Servers or Agents from causing problems in your cluster.

Cloudera Manager supports three levels of TLS security:

- Level 1 (Good): Encrypted communications between the Server and Agents only; no authentication of Server and Agents. See Configuring TLS Encryption for Cloudera Manager below.
- Level 2 (Better): Encrypted communications and authentication of Server to Agents and users; no authentication of Agents to Server. See Configuring TLS Authentication of Server to Agents and Users below.
- Level 3 (Best): Encrypted communications, authentication of Server to Agents, and authentication of Agents to Server. See Configuring TLS Authentication of Agents to Server below.

Configuring TLS Encryption only for Cloudera Manager

Use the keytool located here to manage the public keys and certificates for the Cloudera Manager Server. Before configuring TLS security for Cloudera Manager, create a keystore, as described in the documentation at the preceding link. For example, you might use a command similar to the following:

```
keytool -genkey -alias jetty -keystore truststore
```

Step 1: Create a Cloudera Manager Server certificate.

- **Warning:**

  You must use an Oracle JDK keytool.

1. Use keytool to generate a certificate for the Cloudera Manager Server. For example:

   ```
   $ keytool -validity 180 -keystore <path-to-keystore> -alias jetty -genkeypair
   -keyalg RSA
   ```

   - The `-validity` option specifies the certificate lifetime in number of days. If no `validity` value is specified, the default value is used. The default varies, but is often 90 days.
   - The `<path-to-keystore>` must be a path to where you want to save the keystore file, and where the Cloudera Manager Server host machine can access.

2. When prompted by keytool, create a password for the keystore. Save the password in a safe place.

3. When prompted by keytool, fill in the answers accurately to the questions to describe you and your company. The most important answer is the CN value for the question "What is your first and last name?" The CN must match the fully-qualified domain name (FQDN) or IP address of the host machine where the Server is running. For example, `cmf.company.com` or `192.168.123.101`.

> **Important:**
>
> For the CN value, be sure to use a FQDN if possible, or a static IP address that will not change. Do not specify an IP address that will change periodically. When agents connect to the server using TLS, they check whether the key uses the same name as the one they are using to connect to the server. If the names do not match, agents do not heartbeat.

Step 2: Enable TLS encryption and specify Server keystore properties.

1. Log into the Cloudera Manager Admin Console.
2. From the **Administration** tab select **Settings**, then go to the **Security** category.
3. Configure the following three TLS settings:

Setting	Description
Use TLS Encryption for Agents	Select this option to enable TLS encryption between the Server and Agents.
Path to TLS Keystore File	Specify the full filesystem path to the keystore file.
Keystore Password	Specify the password for keystore.

4. Click **Save Changes** to save the settings.

Step 3: Enable and configure TLS on the Agent machines.

To enable and configure TLS, you must specify values for the TLS properties in the `/etc/cloudera-scm-agent/config.ini` configuration file on all Agent machines.

1. On the Agent Host machine, open the `/etc/cloudera-scm-agent/config.ini` configuration file:
2. Edit the following property in the `/etc/cloudera-scm-agent/config.ini` configuration file.

Property	Description
use_tls	Specify `1` to enable TLS on the Agent, or `0` (zero) to disable TLS.

3. Repeat these steps on every Agent Machine.

Step 4: Restart the Cloudera Manager Server.

Restart the Cloudera Manager Server with the following command to activate the TLS configuration settings.

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

Important:

To enable TLS security, you must restart the Server.

Step 5: Restart the Cloudera Manager Agents.

On every Agent Host machine, restart the Agent:

```
$ sudo service cloudera-scm-agent restart
```

Step 6: Verify that the Server and Agents are communicating.

In the Cloudera Manager Admin Console, open the **Hosts** page. If the Agents heartbeat successfully, TLS encryption is working properly.

Configuring TLS Authentication of Server to Agents and Users

This is the second highest level of TLS security and requires that you provide a server certificate for the Server that is signed through a chain to a trusted root CA. You must also provide the certificate of the CA (Certificate Authority) that signed the Server's server certificate. If you are not working in a production environment, you can also use a self-signed server certificate.

Note:

If the Server's server certificate or the associated CA certificate is missing or expired, the Agents do not allow communications with the Server.

Step 1: Configure TLS encryption.

If you have not already done so, you must configure TLS encryption to use this second level of security. For instructions, see Configuring TLS Encryption for Cloudera Manager.

Step 2: Provide the Server's server certificate and CA certificate.

1. If you already have the Server's server certificate, and the certificate of the CA (Certificate Authority) that signed the Server's server certificate, you can skip down to Copy the Server's server certificate to the Agents below. Alternatively, if you want to generate your own self-signed server certificate, you can use keytool to generate a public certificate for the Server by typing the following command on the Server host:

   ```
   $ keytool -validity 180 -keystore <path-to-keystore> -alias jetty -genkeypair
   -keyalg RSA
   ```

2. When prompted by keytool, create a password for the keystore. Save the password in a safe place.
3. When prompted by keytool, fill in the answers accurately to the questions to describe you and your company. The most important answer is the CN value for the question "What is your first and last name?" The CN must match the fully-qualified domain name (FQDN) or IP address of the host machine where the Server is running. For example, `cmf.company.com` or `192.168.123.101`.

> **Important:**
>
> For the CN value, be sure to use a FQDN if possible, or a static IP address that will not change. Do not specify an IP address that will change periodically. When agents connect to the server using TLS, they check whether the key uses the same name as the one they are using to connect to the server. If the names do not match, agents do not heartbeat.

4. On the Server machine, run the following command to export the server certificate from your keystore in the binary DER format:

```
$ keytool -exportcert -keystore <path-to-keystore> -alias jetty -file server.der
```

5. Convert the binary DER format to a .pem file that can be used on the Agents by using openssl (available for download here.)

```
$ openssl x509 -out server.pem -in server.der -inform der
```

Step 3: Copy the Server's server .pem file to the Agents.

1. Copy the Server's server .pem file (for example, `server.pem`) to the Agent machine in any directory. For example, copy the .pem file to `/etc/cmf`.
2. On the Agent Host machine, open the `/etc/cloudera-scm-agent/config.ini` configuration file:
3. Edit the following property in the `/etc/cloudera-scm-agent/config.ini` configuration file.

Property	Description
verify_cert_file	Enter the path to the Server's server .pem file. For example, `/etc/cmf/server.pem`.

4. Repeat these steps on every Agent Machine.

Step 4: Restart the Cloudera Manager Agents.

On every Agent Host machine, restart the Agent:

```
$ sudo service cloudera-scm-agent restart
```

Step 5: Verify that the Server and Agents are communicating.

In the Cloudera Manager Admin Console, open the **Hosts** page. If the Agents heartbeat successfully, the Server and Agents are communicating. If not, check the Agent log `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` which shows errors if the connection fails.

Step 6: (Optional) Enable Authentication from Server to Users

This is an optional step in which you can enable TLS authentication from the Server to Cloudera Manager users.

> **Warning:**
>
> Do not enable the **Use TLS Encryption for Admin Console** option as described in the following instructions in this step until after you have completed the previous steps in this procedure. If you enable the **Use TLS Encryption for Admin Console** option before performing the previous steps, you will lose the ability to connect to the Cloudera Manager Server from the Admin Console.

1. Log into the Cloudera Manager Admin Console.
2. From the **Administration** tab select **Settings**, then go to the **Security** category.
3. Select the **Use TLS Encryption for Admin Console** option to enable TLS Authentication between the Cloudera Manager Server and the instance of Cloudera Manager that runs in your browser.
4. Click **Save Changes** to save the settings.
5. Restart the Server.

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

6. Log out and then log in into Cloudera Manager to test the certificate. You may see an warning message to accept the certificate if the root certificate is not installed in your browser.
7. Restart the Cloudera Management Services by clicking the **Services** link and choosing **Restart** on the **Actions** menu for the Cloudera Management Services. Click **Restart** that appears in the next screen to confirm. When you see a **Finished** status, the service has restarted.

Configuring TLS Authentication of Agents to Server

This is the highest level of TLS security and requires you to use openssl to create private keys and public certificates for every Agent on your cluster, and import those Agents' certificates into the Server's truststore.

Step 1: Configure TLS encryption.

If you have not already done so, you must configure TLS encryption to use this third level of security. For instructions, see Configuring TLS Encryption for Cloudera Manager.

Step 2: Configure TLS Authentication of Server to Agents.

If you have not already done so, you must configure TLS Authentication of Server to Agents. For instructions, see Configuring TLS Authentication of Server to Agents.

Step 3. Generate the private key for the Agent using openssl.

1. Run the following openssl command on the agent:

```
$ openssl genrsa -des3 -out agent.key
```

2. Provide a password for the key file. Note it in a safe place.

Step 4: Generate a certificate for the agent.

1. Run the following openssl command.

   ```
   $ openssl req -new -x509 -days 365 -key agent.key -out agent.pem
   ```

   The key is output in a .pem file. In the preceding example, the optional `days` argument results in a certificate that is valid for 365 days.

2. Fill in the answers to the questions about the certificate. Note that the CN must match the host name or IP address of the Agent machine.

Step 5: Create a file that contains the password for the key.

The Agent reads the password from a text file instead of from a command line. The file allows you to use file permissions to protect the password. For example, name the file `agent.pw`.

Step 6: Configure the Agent with its private key and certificate.

1. On the Agent Host machine, open the `/etc/cloudera-scm-agent/config.ini` configuration file:
2. Edit the following properties in the `/etc/cloudera-scm-agent/config.ini` configuration file.

Property	Description
client_key_file	Name of client key file
client_keypw_file	Name of client key pw file
client_cert_file	Name of client certificate file

3. Repeat these steps on every Agent Machine.

Step 7: Import the Agent's certificate into the Server's truststore.

The Server's truststore contains the certificates that are required to authenticate clients. Use the following command to import a certificate called, for example, `agent.pem` into a new truststore called, for example, `truststore`.

```
$ keytool -keystore <path-to-truststore> -import -alias <agent-name> -file agent.pem
```

Step 8: Repeat steps 3 through 7 for every agent in your cluster.

> **Important:**
>
> Each Agent's private key and certificate that you import into the Server's truststore must be unique.

Step 9: Enable Agent authentication and configure the Server to use the new truststore.

1. Log into the Cloudera Manager Admin Console.
2. From the **Administration** tab select **Settings**, then go to the **Security** category.
3. Configure the following three TLS settings:

Setting	Description
Use TLS Authentication of Agents to Server	Select this option to enable TLS Authentication of Agents to the Server.
Path to Truststore	Specify the full filesystem path to the truststore located on the Cloudera Manager Server host.
Truststore Password	Specify the password for the truststore.

4. Click **Save Changes** to save the settings.

Step 10: Restart the Server.

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

Step 11: Restart the Cloudera Manager Agents.

On every Agent Host machine, restart the Agent:

```
$ sudo service cloudera-scm-agent restart
```

Step 12: Verify that the Server and Agents are communicating.

In Cloudera Manager Admin Console, open the **Hosts** page. If the Agents heartbeat successfully, the Server and Agents are communicating. If they are not, you may get an error in the Server, such as a `null CA chain` error. This implies either the truststore doesn't contain the Agent certificate or the Agent isn't presenting the certificate. Double check all of your settings. Check the Server's log to verify whether TLS and Agent validation have been enabled correctly.

Upgrading Cloudera Manager

You can upgrade an existing Cloudera Manager to the latest version of Cloudera Manager. Upgrading preserves existing data and settings, while enabling the use of the new features provided with the latest product versions. To enable new features, some new settings are added, and some additional steps may be required, but nothing is removed.

As of Cloudera Manager 4.6, the former Cloudera Manager Free Edition is now known as Cloudera Standard, and includes a number of features that were previously available only with Cloudera Manager Enterprise Edition. Specifically, service and activity monitoring features are now available, and require databases to be set up for their use. Thus, upon upgrading to Cloudera Manager 4.6, you will be asked for database information for these services. (You will have the option to use the embedded PostGreSQL database for this).

Understanding Upgrades

The process for upgrading to Cloudera Manager varies based on the starting point. The categories of tasks to be completed include the following:

- Install any databases that are newly required for this release. (If you are upgrading a Free Edition installation, you are asked to configure databases for the monitoring features that are now part of Cloudera Standard).
- Upgrade the Cloudera Manager server.
- Upgrade the hosts in the cluster.

Before Upgrading

- The Cloudera Manager Server must have SSH access to the cluster hosts and you must log in using a root account or an account that has password-less sudo permission. See Requirements for Cloudera Manager for more information.
- Ensure there are no running commands. Use the Admin Console's main navigation bar to check for any running commands. You can either wait for commands to complete or abort any running commands. For more information on viewing and aborting running commands, see Viewing Running and Recent Commands.
- Ensure you have completed any required process for preparing databases, as described in Database Considerations for Cloudera Manager Upgrades.

During the Upgrade

During the upgrade process, the following changes occur:

- The database schemas are modified for any databases storing information for Cloudera Manager Server, Activity Monitor, Service Monitor, Report Manager, and Host Monitor.
- Configuration information is reorganized.

After Upgrading

After completing an upgrade to the latest Cloudera Manager (4.6 or later), the following is true:

- You have re-deployed client configurations to ensure client services have the most current configuration.

- Required databases are established to store information for Cloudera Manager Server, Hive Metastore, Activity Monitor, Service Monitor, Report Manager, and Host Monitor.
- The database schemas reflect the current version.
- The Host Monitor service is added and active.
- The Cloudera Manager Server and all supporting services, such as the Activity Monitor, Service Monitor, Report Manager, and Host Monitor are updated.

Upgrade Paths

In some cases, completing an upgrade requires changes to your environment, and in other cases, elements are already in place. For example, if you are upgrading your environment from 3.7 to 4.5 or later, you must add Host Monitor, but if you are upgrading from 4.0 beta or 4.0 GA to 4.6, this is not required, as Host Monitor is included in 4.0. The specific steps required vary based on the path taken.

To upgrade from a version older than Cloudera Manager 3.7, begin by upgrading toCloudera Manager 3.7, and then proceed to upgrade from there.

- **Warning:**

  Cloudera Manager 4.5 or later works with CDH3 and CDH4, but does not work with CDH4.0 beta. You must upgrade any installations of CDH4.0 beta.

Begin the upgrade process by evaluating Database Considerations for Cloudera Manager Upgrades.

Upgrading CDH

Cloudera Manager 4 can manage both CDH3 and CDH4, so upgrading existing CDH3 or CDH4 installations is not required. However, to get the benefits of the most current CDH4 features, you may want to upgrade CDH. See the following topics for more information on upgrading CDH:

- **Important:** Cloudera Manager version 3.x and CDH3 have reached End of Maintenance (EOM) as of June 20, 2013. Cloudera will not support or provide patches for any of the Cloudera Manager version 3.x and CDH3 releases. Even though Cloudera Manager 4.x will continue to support CDH3, it is strongly recommended that you upgrade to CDH4. See Upgrading existing installations of CDH3 to CDH4 for more details.

- Upgrading to the Latest Version of CDH4 in a Cloudera Manager Deployment - Follow this path to upgrade existing installations of CDH4 to the latest version of CDH4.
- Upgrading CDH3 to CDH4 in a Cloudera Managed Deployment - Follow this path to upgrade existing installations of CDH3 to the latest version of CDH4. You can also install Impala when you upgrade to CDH4 version 4.1.2 or later.

Database Considerations for Cloudera Manager Upgrades

Cloudera Manager uses databases to store information about system configurations and tasks. Before upgrading, complete the pre-upgrade database tasks that apply in your environment.

> **Note:** Cloudera Manager 4.5 added support for Hive, which includes a new role type called the Hive Metastore Server. This role manages the metastore process when Hive is configured with a remote metastore.
>
> When upgrading from a previous CDH version, Cloudera Manager automatically creates new Hive service(s) to capture the previous implicit Hive dependency from Hue and Impala. Your previous services will continue to function without impact.
>
> Note that if Hue was using a Hive metastore of type Derby, then the newly created Hive service will also use Derby. But since Derby does not allow concurrent connections, although Hue will continue to work, the new Hive Metastore Server will fail to run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve the set of cluster functionality as it was before upgrade. Cloudera discourages the use of a Derby metastore due to its limitations. You should consider switching to a different supported database type (PostgreSQL, MySQL, Oracle).

After you have completed these steps, the upgrade processes automatically complete any additional updates to database schemas and service data stored. You do not need to complete any data migration.

Back up Databases

Before beginning the upgrade process, shut down the services that are using databases. This includes Cloudera Manager Server, Activity Monitor, Service Monitor, Report Manager, Host Monitor, Cloudera Navigator, and Hive Metastore. Cloudera recommends that you then back up all databases. This is especially important if you are upgrading from 3.7.x and there is any possibility you may want to revert to using 3.7.x. For information on backing up databases:

- For MySQL, see Backing up the MySQL Database.
- For PostgreSQL, see Backing up the PostgreSQL Database.
- For Oracle, work with your database administrator to ensure databases are properly backed up.

If any additional database will be required as a result of the upgrade, complete any required preparatory work to install and configure those databases. For example, Cloudera Manager 4.0 offers a Host Monitoring service that requires a database. To enable the Host Monitoring service, you must install a database. The upgrade instructions assume all required databases have been prepared. For more information on using databases, see Installing and Configuring Databases.

Modify Databases to Support UTF-8

Cloudera Manager 4.0 adds support for UTF-8 character sets. Update any existing databases in your environment that are not configured to support UTF-8.

Modifying MySQL to Support UTF-8

To modify a MySQL database to support UTF-8, the default character set must be changed and then you must restart the mysql service. Use the following commands to complete these tasks:

```
mysql> alter database default character set utf8;
mysql> quit
$ sudo service mysql restart
```

Modifying PostgreSQL to Support UTF-8

There is no single command available to modify an existing PostgreSQL database to support UTF-8. As a result, you must complete the following process:

1. Use `pg_dump` to export the database to a file. This creates a backup of the database that you will import into a new, empty database that supports UTF-8.
2. Drop the existing database. This deletes the existing database.
3. Create a new database that supports Unicode encoding and that has the same name as the old database. Use a command of the following form, replacing the database name and user name with values that match your environment:

   ```
   CREATE DATABASE scm_database WITH OWNER scm_user ENCODING 'UTF8'
   ```

4. Review the contents of the exported database for non-standard characters. If you find unexpected characters, modify these so the database backup file contains the expected data.
5. Import the database backup to the newly created database.

Modifying Oracle to Support UTF-8

Work with your Oracle database administrator to ensure any Oracle databases support UTF-8.

Modify Databases to Support Appropriate Maximum Connections

Check existing databases configurations to ensure the proper maximum number of connections is supported. Update the maximum configuration values, as required.

Modifying the Maximum Number of MySQL Connections

Allow 100 maximum connections for each database and then add 50 extra connections. For example, for two databases set the maximum connections to 250. If you store seven databases on one host (the databases for Activity Monitor, Service Monitor, Report Manager, Host Monitor, Cloudera Manager Server, Cloudera Navigator, and Hive Metastore), set the maximum connections to 750.

Modifying the Maximum Number of PostgreSQL Connections

Update the `max_connection` parameter in the `/etc/postgresql.conf` file.

You may have to increase the system resources available to PostgreSQL, as described at http://www.postgresql.org/docs/9.1/static/kernel-resources.html.

Modifying the Maximum Number of Oracle Connections

Work with your Oracle database administrator to ensure appropriate values are applied for your Oracle database settings. You must determine the number of connections, transactions, and sessions to be allowed. Allow 100 maximum connections for each database and then add 50 extra connections. For example, for two databases set the maximum connections to 250. If you store seven databases on one host (the databases for Activity Monitor, Service Monitor, Report Manager, Host Monitor, Cloudera Manager Server, Cloudera Navigator, and Hive Metastore), set the maximum connections to 750.

From the maximum number of connections, you can determine the number of anticipated sessions using the following formula:

```
sessions = (1.1 * maximum_connections) + 5
```

For example, if a host has two databases, you anticipate 250 maximum connections. If you anticipate a maximum of 250 connections, plan for 280 sessions.

Once you know the number of sessions, you can determine the number of anticipated transactions using the following formula:

```
transactions = 1.1 * sessions
```

Continuing with the previous example, if you anticipate 280 sessions, you can plan for 308 transactions.

Work with your Oracle database administrator to apply these derived values to your system.

Using the sample values above, Oracle attributes would be set as follows:

```
alter system set processes=250;
alter system set transactions=308;
alter system set sessions=280;
```

Next Steps

After you have completed any required database preparatory tasks, continue to the upgrade path that is appropriate for your environment. Supported paths include:

- Upgrade Cloudera Manager 3.7.x to the Latest Cloudera Manager on page 33
- Upgrade Cloudera Manager 4 to the Latest Cloudera Manager on page 41
- Upgrade from Cloudera Standard to Cloudera Enterprise on page 50

Upgrade Cloudera Manager 3.7.x to the Latest Cloudera Manager

> **Important:**
>
> Cloudera Manager version 3 has reached End of Maintenance (EOM) as of June 20, 2013 . Cloudera will not support or provide patches for any of the Cloudera Manager version 3 releases.

Upgrading from Cloudera Manager 3.7 (either Free or Enterprise Edition) to the latest version of Cloudera Manager involves upgrading Cloudera Manager Server packages, and updating the Cloudera manager agents on all cluster hosts. If you are upgrading a Free Edition version, the upgrade also includes adding databases for the management services supported as of Cloudera Manager 4.6.

> **Note:** As of Cloudera Manager 4.6, the former Cloudera Manager Free Edition is now known as Cloudera Standard, and includes a number of features that were previously available only with Cloudera Manager Enterprise Edition. Specifically, service and activity monitoring features are now available, and require databases to be set up for their use. Thus, upon upgrading to Cloudera Manager 4.6, you will be asked for database information for these services. (You will have the option to use the embedded PostGreSQL database for this).

> **Note:** Cloudera Manager 4.x can manage CDH3 and CDH4, but cannot manage CDH4.0 beta. When you upgrade to Cloudera Manager 4.x, you must upgrade any existing installations of CDH4.0 beta, as well.

To complete the upgrade, you stop the Cloudera Management Service, upgrade the packages (and database tables, if necessary), and then start the Cloudera Management Service again. After upgrading Cloudera Manager you may also want to upgrade CDH or add Impala.

Important:

Cloudera Manager 4.5 added support for Hive, which includes a new role type called the Hive Metastore Server. This role manages the metastore process when Hive is configured with a remote metastore.

When upgrading from a version of Cloudera Manager prior to 4.5, Cloudera Manager automatically creates new Hive service(s) to capture the previous implicit Hive dependency from Hue. Your previous services will continue to function without impact.

Note that if Hue was using a Hive metastore of type Derby, then the newly created Hive service will also use Derby. But since Derby does not allow concurrent connections, Hue will continue to work, but the new Hive Metastore Server will fail to run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve the set of cluster functionality as it was before upgrade. Cloudera discourages the use of a Derby metastore due to its limitations. You should consider switching to a different supported database type (PostgreSQL, MySQL, Oracle).

Cloudera Manager provides a Hive configuration option to bypass the Hive Metastore server. When this configuration is enabled, Hive clients, Hue, and Impala connect directly to the Hive Metastore Database. Prior to Cloudera Manager 4.5, Hue and Impala talked directly to the Hive Metastore Database, so the Bypass mode is enabled by default when upgrading to Cloudera Manager 4.5 or later. This is to ensure the upgrade doesn't disrupt your existing setup. You should plan to disable the Bypass Hive Metastore Server mode, especially when using CDH 4.2 or later. Using the Hive Metastore Server is the recommended configuration. After changing this configuration, you must re-deploy your client configurations, restart Hive, and restart any Hue or Impala services configured to use that Hive.

Cloudera Manager 4.5 or later also supports Hive Server2 with CDH4.2. Hive Server2 is not added by default, but can be added as a new role under the Hive service (see Adding Role Instances).

If you were using the Enterprise Edition of Cloudera Manager, Cloudera Manager 4 adds an optional Host Monitor, which requires an additional database. If you intend to deploy Host Monitor or any other additional agents, you must establish a database for them. For information on establishing databases for agents such as Host Monitor, see Installing and Configuring Databases.

Important:

You will need to restart your clusters after you have finished the Cloudera Manager upgrade.

Summary: What You are Going to Do

Upgrading from Cloudera Manager 3.7.x to Cloudera Manager 4.6 involves the following broad steps:

Step 1. Stop the Cloudera Management Service – Stop the Cloudera Management Service.

Step 2. Upgrade Cloudera Manager Server – Stop Cloudera Manager services, copy files to the Cloudera Manager server, upgrade the server, and restart the server.

Step 3. Upgrade the Cluster Hosts – Use the upgrade wizard to upgrade hosts in the cluster.

Step 4. Verify the Upgrade – You can choose to check the versions of installed components.

Step 5. Add Hive Gateway Roles – Add Hive gateway roles on Hive client hosts to ensure Hive client configurations are deployed on those hosts.

Step 6. Restart the Cloudera Management Service – Restart the Cloudera Manager Management Service. If you have ZooKeeper installed, you will also need to restart the ZooKeeper service so that the ZooKeeper health checks will succeed.

Step 7. Deploy Updated Client Configurations – Update client configurations to ensure clients operate as expected with the upgraded systems.

Step 8. Restart Your Cluster(s) – You must restart your cluster to ensure compatibility with the updated JDK.

Step 9. (Optional) Upgrade CDH – You may choose to upgrade CDH installations. Cloudera Manager 4 can manage both CDH 3 and CDH 4.

Before beginning the upgrade, follow the guidelines described in Database Considerations for Cloudera Manager Upgrades.

After completing the upgrade from Cloudera Manager 3.7.x to Cloudera Manager 4.6, as described in this topic, all required updates to database schemas and service data is completed automatically. You do not need to complete any additional database updates or data migration.

> **Warning:**
>
> Cloudera Manager 4.x can manage CDH3 and CDH4, but cannot manage CDH4.0 beta. If you upgrade to Cloudera Manager 4.x, you must upgrade any existing installations of CDH4.0 beta, as well.

Step 1. Stop the Cloudera Management Service

The Cloudera Manager Service must be stopped before upgrades can occur.

To stop the Cloudera Management Service

1. Click the **Services** tab in Cloudera Manager Admin Console.
2. Choose **Stop** on the **Actions** menu for the Cloudera Management Services.

Step 2. Upgrade Cloudera Manager Server

This process involves stopping running Cloudera Manager service, downloading and applying updates to Cloudera Manager, and restarting the Cloudera Manager service. Valid licenses from Cloudera Manager 3.7.x continue to work with Cloudera Manager 4.

You can use package management software to download and apply updates from Cloudera's software repository. The default name of the repo file is `cloudera-manager`. This name is also typically in square brackets on the first line of the Cloudera Manager repo file. For example, you could view the contents of the repo file, including the repo name in brackets. This file might be at `/etc/yum.repos.d/cloudera-manager.repo` and its contents could be viewed using the `more` command as follows:

```
[user@system yum.repos.d]$ more cloudera-manager.repo
[cloudera-manager]
...
```

The location of the repo files varies by operating system and package management solution.

- For yum the repo file is at `/etc/yum.repos.d/cloudera-manager.repo`.
- For zypper the repo file is at `/etc/zypp/repos.d/cloudera-manager.repo`.

Find Cloudera's repo file for your distribution by starting at `http://archive.cloudera.com/cm4/` and navigating to the directory that matches your operating system. For example, for Red Hat 6, you would navigate to `http://archive.cloudera.com/cm4/redhat/6/x86_64/cm/`. Within that directory, find the repo file that contains information including the repository's base URL and gpgkey. In the preceding example, the contents of the `cloudera-manager.repo` file might appear as follows:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 4, on RedHat or CentOS 5 x86_64
```

```
name=Cloudera Manager
baseurl=http://archive.cloudera.com/cm4/redhat/5/x86_64/cm/4/
gpgkey = http://archive.cloudera.com/cm4/redhat/5/x86_64/cm/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

Copy this repo file to the configuration location for the package management software for your system. Continuing with the preceding example, on Red Hat 6, you would copy the `cloudera-manager.repo` file to `/etc/yum.repos.d/`.

Before beginning the upgrade process, it can be best to clean all `yum`'s cache directories using the command `yum clean all`. Doing so ensures that you download and install the latest versions of the packages. If your system is not up to date, and any underlying system components need to be upgraded before this `yum update` can succeed, yum will tell you what those are.

To upgrade to the new server

1. Stop the server on the 3.7.x Server host:

   ```
   $ sudo service cloudera-scm-server stop
   ```

2. If you are using the embedded PostgreSQL database, stop `cloudera-manager-server-db` on the host on which it is running:

   ```
   $ sudo service cloudera-scm-server-db stop
   ```

3. Install the new version of the server. You can run commands on the Cloudera Manager Server host to update only the Cloudera Manager components.

   For a Red Hat system:

   To upgrade from Cloudera's repository run the following commands on the Cloudera Manager Server host:

   ```
   $ sudo yum clean all
   $ sudo yum update 'cloudera-*'
   ```

   On a SLES system:

   Use the following commands to clean the cached repository information and update only the Cloudera Manager components:

   ```
   $ sudo zypper clean --all
   $ sudo zypper up -r http://archive.cloudera.com/cm4/sles/11/x86_64/cm/4/
   ```

   At the end of this process you should have the 4.5 versions of the following packages installed on the host that will become the Cloudera Manager Server host. For example,

   ```
   $ rpm -qa 'cloudera-manager-*'
   cloudera-manager-server-4.6.0-1.cm460.p0.99.x86_64
   cloudera-manager-agent-4.6.0-1.cm460.p0.99.x86_64
   cloudera-manager-daemons-4.6.0-1.cm460.p0.99.x86_64
   ```

   You may also see additional packages for plugins, depending on what was previously installed on the Server host.

4. Start the server. If you are using the embedded PostgreSQL database, start `cloudera-scm-server-db` on the Cloudera Manager Sever host:

   ```
   $ sudo service cloudera-scm-server-db start
   ```

You will see it upgrade and create tables and databases. On the Cloudera Manager Server host (the system on which you installed the `cloudera-manager-server` package) do the following:

```
$ sudo service cloudera-scm-server start
```

You should see the following:

```
Starting cloudera-scm-server:                                      [  OK  ]
```

Note:

If you have problems starting the server, such as database permissions problems, you can use the server's log `/var/log/cloudera-scm-server/cloudera-scm-server.log` to troubleshoot the problem.

Step 3. Upgrade the Cluster Hosts

After updating Cloudera Manager, connect to Cloudera Manager and use the wizard to continue the upgrade process. In this part of the process, the Cloudera Manager agents and their databases are updated. The Host Monitor role is a new addition for Cloudera Manager 4, so upgrading includes adding this role and its supporting database. If you are adding new agents, such as the Host Monitor, you must have a database available to support these roles. For more information, see Installing and Configuring Databases.

Important:

All hosts in the cluster must have access to the Internet if you plan to use `archive.cloudera.com` as the source for installation files. If you do not have Internet access, create a custom repository.

1. Log in to the Cloudera Manager Admin Console. If you have just restarted the Cloudera Manager server, you may need to log in again.
2. On the Welcome screen, select whether you want to:
   - Install Cloudera Standard,
   - Try the Cloudera Enterprise with a 60-day trial license, or
   - Install a license you have purchased for Cloudera Enterprise.
3. After you upload the Cloudera Manager license, or if you have elected to use a Trial license, restart the Cloudera Manager server.

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

   - As the Cloudera Manager server restarts, the UI indicates its progress, and presents the login page when the restart has completed.
4. Click **Continue** to proceed to the Upgrade cluster hosts screen.
5. On the Upgrade cluster hosts screen, click **Start Upgrade** to upgrade the existing managed hosts. Click **Skip Host Upgrades** to skip this step.
6. Select the release of the Cloudera Manager Agent to install. Normally, this will be the **Matched Release for this Cloudera Manager Server**. However, if you used a custom repository for the Cloudera Manager server, select **Custom Repository** and provide the required information

   Click **Continue** to proceed.
7. Provide credentials for authenticating with hosts.

a. Select **root** or enter the user name for an account that has password-less sudo permissions.
b. Select an authentication method.

   - If you choose to use password authentication, enter and confirm the password.
   - If you choose to use public-key authentication provide a passphrase and path to the required key files.
   - You can choose to specify an alternate SSH port. The default value is 22.
   - You can specify the maximum number of host installations to run at once. The default value is 10.

8. Click **Start Installation** to install and start Cloudera Manager Agents. The status of installation on each host is displayed on the page that appears after you click **Start Installation**. You can also click the **Details** link for individual hosts to view detailed information about the installation and error messages if installation fails on any hosts.

   Note:

   If you click the **Abort Installation** button while installation is in progress, it will halt any pending or in-progress installations and roll back any in-progress installations to a clean state. The **Abort Installation** button does not affect host installations that have already completed successfully or already failed.

   If installation fails on a host, you can click the **Retry** link next to the failed host to try installation on that host again. To retry installation on all failed hosts, click **Retry Failed Hosts** at the bottom of the screen.

9. When the **Continue** button appears at the bottom of the screen, the installation process is complete. If the installation has completed successfully on some hosts but failed on others, you can click **Continue** if you want to skip installation on the failed hosts and continue to the next screen to start installing the Cloudera Management services on the successful hosts.
10. On the next screen, click **Continue** to install the Cloudera Management services.
11. The Host Inspector runs to validate your installation. This should show your currently installed components as CDH3, with CDH4 components shown as **Not installed**. Note that the Version will be shown as **Unavailable** for all components.
12. Select the Cloudera Management Service roles you want to install. The wizard evaluates the hardware configurations of the cluster hosts to recommend the best machines for each role. The Host Monitor is a new role introduced in Cloudera Manager 4.1. Navigator is a new, independently-licensed feature introduced with Cloudera Manager 4.5.

    Important:

    For best performance, make sure the Host Monitor role is assigned to the host on which you installed the corresponding databases. For example, if you created the Host Monitor database on `myhost1`, then you should assign the Activity Monitor role to `myhost1`. The JDBC connector **must** be installed and configured on any machine to which any of these roles is assigned.

    Click **Continue** to proceed.

13. On the Database Setup page, enter any required information for Host Monitor and Navigator databases.

    Important:

    The value you enter as the database hostname **must** match the value you entered for the hostname (if any) when you created the database (see Installing and Configuring Databases).

    a. Enter the fully-qualified domain name for the server that is hosting the database in **Database Host Name**.
    b. Select the proper database type from the choices provided in **Database Type**.
    c. Enter the name you specified when you created the database in **Database Name**.

d. Enter the user name you specified when you created the database in **Username**.
e. Enter the password you specified when you created the database in **Password**.

> **Note:**
>
> Problems may occur if a database with a blank password is used.

14. Click **Test Connection** to confirm that Cloudera Manager can communicate with the databases using the information you have supplied. This transaction takes two heartbeats to complete (about 30 seconds with the default heartbeat interval). If the test succeeds in all cases, click **Continue**; otherwise check and correct the information you have provided for the databases and then try the test again.
15. Review the configuration changes to be applied during the upgrade and click **Accept**.
16. On the next page, select the hosts where the Hive Metastore Server role should be installed. The Hive service is now managed by Cloudera Manager; you must select the host for the Hive MetaStore Server. You should assign the Hive Metastore server to a single host. Click **Continue** to proceed.
17. Review the configuration values for your Hive roles, and click **Accept** to continue.

> **Note:** If Hue is using a Hive metastore of type Derby (the default), then the newly created Hive service will also use Derby. However, since Derby does not allow concurrent connection, Hue will continue to work but the new Hive Metastore Server will fail to start. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve the cluster functionality that existed before the upgrade. Hive's Bypass Metastore Server mode is enabled by default when upgrading to Cloudera Manager 4.5. This is to ensure the upgrade doesn't disrupt your existing setup. You should plan to disable the Bypass Hive Metastore Server mode, especially when using CDH 4.2 or later. Using the Hive Metastore Server is the recommended configuration. After changing this configuration, you must re-deploy your client configurations, restart Hive, and restart any Hue or Impala services configured to use Hive.

18. You are now taken to the Hive service **Instances** page: The Hive metastore server will be stopped.
19. Under the **Services** tab, click the **All Services** link to go to the service overview page. All the services except for Hive and the Cloudera Management Service should now be running.

Step 4. Verify the Upgrade

You can use the host inspector to verify the upgrade completed.

To verify the upgrade has completed as expected

1. Connect to the Cloudera Manager Admin Console.
2. Click the **Hosts** tab.
3. Click **Host Inspector**.
4. Click **Show Inspector Results**. All results from the host inspector process are displayed including the currently installed versions. If this includes listings of current component versions, the installation completed as expected.

Step 5. Add Hive Gateway Roles to Hosts

1. Add Hive Gateway roles to any hosts where Hive clients should run.
2. In the Cloudera Manager Admin console, pull down the **Services** tab and select the Hive service.
3. Go to the **Instances** tab, and click the **Add** button. This opens the **Add Role Instances** page.
4. Select the hosts on which you want a Hive Gateway role to run. This will ensure that the Hive client configurations are deployed on these hosts.

Step 6. Restart the Cloudera Management Service

The Cloudera Management Services are not started automatically after an upgrade — you must restart them.

To start the Cloudera Management Service

1. Click the **Services** tab and select **All Services** in the Cloudera Manager Admin Console.
2. Choose **Start** on the **Actions** menu for the Cloudera Management Services. If you are running more than one cluster, you should do this for each one.

Adding the Cloudera Navigator Role

If you have upgraded to Cloudera Enterprise or are running the 60-day Trial and want to try Cloudera Navigator, you must add it as a role under the management service.

1. From the Services page, select the Management Service.
2. Go to the **Instances** tab, and click the **Add** button.
3. In the table presented, scroll to the end and select the host where you want the Navigator Server role to be hosted, and click **Continue**.
4. Because Cloudera Navigator is separately licensed, you are presented with a license statement. Click **Accept** to enable the trial license for this feature.
5. Enter the credentials for the database to be used by the Navigator Server. Assuming you have not set up an external database, you can use the Embedded Database for this. Click **Test Connection** to verify connectivity to the Database, the click **Continue**.
6. Review and accept any configuration changes (typically there are none). Click **Accept**. This returns you to the Instances page.
7. The Navigator Server role is added but not started. To start the role:
   a. Click the checkbox next to the role.
   b. From the **Actions for Selected** menu, click **Start**, and confirm that you want to start the role.

Step 7. Deploy Updated Client Configurations

During upgrades between major versions, resource locations may change. To ensure clients have current information about resources, update client configuration as described in Deploying Client Configuration Files.

Step 8. Restart Your Cluster(s)

From the **Actions** menu for each cluster, click **Restart**.

Step 9. (Optional) Upgrade CDH

Cloudera Manager 4.x can manage both CDH3 and CDH4, so upgrading existing installations is not required. However, to get the benefits of CDH4, you may want to upgrade to the latest version. If you are using CDH4.0 beta, you must upgrade to a newer version of CDH4.

See the following topics for more information on upgrading CDH:

- Upgrading to the Latest Version of CDH4 in a Cloudera Managed Deployment - Follow this path to upgrade existing installations of CDH4 to the latest version of CDH4.
- Upgrading CDH3 to CDH4 in a Cloudera Managed Deployment - Follow this path to upgrade existing installations of CDH3 to the latest version of CDH4. You can also install Impala when you upgrade to CDH4 version 4.1.2 or later.

Upgrade Cloudera Manager 4 to the Latest Cloudera Manager

Upgrading from an earlier version of Cloudera Manager 4 (either Free or Enterprise Edition) to the latest version of Cloudera Manager is a relatively simple process, that primarily involves upgrading Cloudera Manager Server packages. This process applies to upgrading Cloudera Manager 4.0.x, 4.1.x, 4.5.x and 4.6.x to the latest available version of Cloudera Manager.

> **Note:** As of Cloudera Manager 4.6, the former Cloudera Manager Free Edition is now known as Cloudera Standard, and includes a number of features that were previously available only with Cloudera Manager Enterprise Edition. Specifically, service and activity monitoring features are now available, and require databases to be set up for their use. Thus, upon upgrading to Cloudera Manager 4.6, you will be asked for database information for these services. (You will have the option to use the embedded PostgreSQL database for this).

To complete the upgrade, you stop the Cloudera Management Service, upgrade the packages (and database tables, if necessary), and then start the Cloudera Management Service again. This should not affect your CDH installation, although you may need to stop some dependent services. After upgrading Cloudera Manager you may also want to upgrade CDH or add Cloudera Impala or Cloudera Search.

It is possible to complete the following upgrade without shutting down the Hadoop services. Hadoop daemons can continue running, unaffected, while Cloudera Manager is upgraded.

> **Important:**
>
> Cloudera Manager 4.5 added support for Hive, which includes a new role type called the Hive Metastore Server. This role manages the metastore process when Hive is configured with a remote metastore.
>
> When upgrading from a version of Cloudera Manager prior to 4.5, Cloudera Manager automatically creates new Hive service(s) to capture the previous implicit Hive dependency from Hue and Impala. Your previous services will continue to function without impact.
>
> Note that if Hue was using a Hive metastore of type Derby, then the newly created Hive service will also use Derby. But since Derby does not allow concurrent connections, Hue will continue to work, but the new Hive Metastore Server will fail to run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve the set of cluster functionality as it was before upgrade. Cloudera discourages the use of a Derby metastore due to its limitations. You should consider switching to a different supported database type (PostgreSQL, MySQL, Oracle).
>
> Cloudera Manager provides a Hive configuration option to bypass the Hive Metastore server. When this configuration is enabled, Hive clients, Hue, and Impala connect directly to the Hive Metastore Database. Prior to Cloudera Manager 4.5, Hue and Impala talked directly to the Hive Metastore Database, so the Bypass mode is enabled by default when upgrading to Cloudera Manager 4.5 or later. This is to ensure the upgrade doesn't disrupt your existing setup. You should plan to disable the Bypass Hive Metastore Server mode, especially when using CDH 4.2 or later. Using the Hive Metastore Server is the recommended configuration. After changing this configuration, you must re-deploy your client configurations, restart Hive, and restart any Hue or Impala services configured to use that Hive.
>
> Cloudera Manager 4.5 or later also supports Hive Server2 with CDH4.2. Hive Server2 is not added by default, but can be added as a new role under the Hive service (see Adding Role Instances).

Summary: What You are Going to Do

Upgrading from a version of Cloudera Manager 4 to the latest version of Cloudera Manager involves the following broad steps:

Step 1. Stop Selected Services

Step 2. Upgrade the Cloudera Manager Server Server

Step 3. Start the Server

Step 4. Upgrade the Cluster Hosts

Step 5. Verify the Upgrade Succeeded

Step 6. Add Hive Gateway roles

Step 7. Restart the Services You Stopped

Step 8. Deploy Updated Client Configurations

Step 9. (Optional) Upgrade CDH

Step 1. Stop Selected Services as Needed

Stop the Cloudera Management Service, if it is running, and stop any services that depend on the Hive metastore.

If you are upgrading from the Enterprise Edition, you must stop the Cloudera Management service before upgrades can occur.

To stop the Cloudera Management Service:

1. From the **Services** tab select **All Services** in the Cloudera Manager Admin Console.
2. Choose **Stop** on the **Actions** menu for the Cloudera Management Services.

If you are upgrading from Cloudera Manager 4.5 to a newer version, *and* you are using the embedded PostgreSQL database, you must stop the services that have a dependency on the Hive Metastore (Hive, Hue, and Impala). You will not be able to stop the Cloudera Manager server's database while these services are running.

- Choose **Stop** on the **Actions** menus for the Hive and Hue services. Do the same for Impala if you have it running.

Step 2. Upgrade the Cloudera Manager Server Software

In this step, you upgrade the Cloudera Manager Server packages to the latest version. The Agents' packages will be updated in Step 4. Deploy the Upgraded Software.

1. Stop the server and the server's database on the Cloudera Manager Server host using the following commands:

   ```
   $ sudo service cloudera-scm-server stop
   ```

2. **If you are using the embedded PostgreSQL database for Cloudera Manager**, stop the database on the Cloudera Manager Server host:

   ```
   $ sudo service cloudera-scm-server-db stop
   ```

   If you are not using the embedded database, you should skip this step.
3. Install the new version of the server. To install the new version, you can upgrade from Cloudera's repository at `http://archive.cloudera.com/cm4/`. Alternately, you can create your own repository, as described in

Appendix A - Understanding Custom Installation Solutions. Creating your own repository is necessary if you are upgrading a cluster that does not have access to the Internet.

a. Find Cloudera's repo file for your distribution by starting at `http://archive.cloudera.com/cm4/` and navigating to the directory that matches your operating system. For example, for Red Hat or CentOS 6, you would navigate to `http://archive.cloudera.com/cm4/redhat/6/x86_64/cm/`. Within that directory, find the repo file that contains information including the repository's base URL and gpgkey. In the preceding example, the contents of the `cloudera-manager.repo` file might appear as follows:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 4, on RedHat or CentOS 5 x86_64
name=Cloudera Manager
baseurl=http://archive.cloudera.com/cm4/redhat/5/x86_64/cm/4/
gpgkey = http://archive.cloudera.com/cm4/redhat/5/x86_64/cm/RPM-GPG-KEY-cloudera

gpgcheck = 1
```

For Ubuntu or Debian systems, the repo file can be found by navigating to the appropriate directory, for example, `http://archive.cloudera.com/cm4/debian/squeeze/amd64/cm`. The repo file, in this case, `cloudera.list`, may appear as follows:

```

# Packages for Cloudera's Distribution for Hadoop, Version 4, on Debian 6.0
x86_64
deb http://archive.cloudera.com/cm4/debian/squeeze/amd64/cm squeeze-cm4 contrib
deb-src http://archive.cloudera.com/cm4/debian/squeeze/amd64/cm squeeze-cm4
contrib
```

Copy this repo file to the configuration location for the package management software for your system. For example, with Red Hat 6, you would copy the `cloudera-manager.repo` file to `/etc/yum.repos.d/`. For SLES, you would copy the `cloudera-manager.repo` file to `/etc/zypp/repos.d/`. For Ubuntu/Debian, you would copy the `cloudera.list` file, to `/etc/apt/sources.list.d/`.

b. After verifying that you have the correct repo file, run the following commands:

<table>
<tr><th>Operating System</th><th>Commands</th></tr>
<tr><td><b>RHEL</b></td><td>

```
$ sudo yum clean all
$ sudo yum update 'cloudera-*'
```

Note:

- `yum clean all` cleans up `yum`'s cache directories, ensuring that you download and install the latest versions of the packages.
- If your system is not up to date, and any underlying system components need to be upgraded before this `yum update` can succeed, yum will tell you what those are.

</td></tr>
<tr><td><b>SLES</b></td><td>

```
$ sudo zypper clean --all
$ sudo zypper up -r
http://archive.cloudera.com/cm4/sles/11/x86_64/cm/4/
```

To download from your own repository:

```
$ sudo zypper clean --all
$ sudo zypper rr cm
$ sudo zypper ar -t rpm-md
http://myhost.example.com/path_to_cm_repo/ cm
$ sudo zypper up -r http://myhost.example.com/path_to_cm_repo
```

</td></tr>
<tr><td><b>Ubuntu or Debian</b></td><td>

Use the following commands to clean cached repository information and update Cloudera Manager components:

```
$ sudo apt-get clean
$ sudo apt-get update
$ sudo apt-get install cloudera-manager-server
cloudera-manager-agent cloudera-manager-daemons
```

As this process proceeds, you may be prompted concerning your configuration file version:

```
Configuration file `/etc/cloudera-scm-agent/config.ini'
==> Modified (by you or by a script) since installation.
==> Package distributor has shipped an updated version.
What would you like to do about it ? Your options are:
Y or I : install the package maintainer's version
N or O : keep your currently-installed version
D : show the differences between the versions
Z : start a shell to examine the situation
The default action is to keep your current version.
```

You will receive a similar prompt for `/etc/cloudera-scm-server/db.properties`. *Answer* **N** *to both these prompts.*

</td></tr>
</table>

At the end of this process you should have the following packages, corresponding to the version of Cloudera Manager you installed, on the host that will become the Cloudera Manager Server host. For example, for CentOS,

```
$ rpm -qa 'cloudera-manager-*'
cloudera-manager-agent-4.6.1-1.cm461.p0.164.x86_64
cloudera-manager-daemons-4.6.1-1.cm461.p0.164.x86_64
cloudera-manager-server-4.6.1-1.cm461.p0.164.x86_64
```

For Ubuntu or Debian, you should have packages similar to those shown below.

```
~# dpkg-query -l 'cloudera-manager-*'
Desired=Unknown/Install/Remove/Purge/Hold
| Status=Not/Inst/Conf-files/Unpacked/halF-conf/Half-inst/trig-aWait/Trig-pend
|/ Err?=(none)/Reinst-required (Status,Err: uppercase=bad)
||/ Name                                                       Version
                                        Description
+++-===============-===================-===========================================
ii  cloudera-manager-daemons                  4.6.1-1.cm461.p0.175~squeeze-cm4.6.1
              Provides daemons for monitoring Hadoop and related tools.
ii   cloudera-manager-repository                 4.0
                 Cloudera Manager
ii  cloudera-manager-server                   4.6.1-1.cm461.p0.175~squeeze-cm4.6.1
              The Cloudera Manager Server
ii  cloudera-manager-server-db                4.6.1-1.cm461.p0.175~squeeze-cm4.6.1
              This package configures an "embedded" PostgreSQL server,running as
user cloudera-scm on port 7432.
```

You may also see an entry for the `cloudera-manager-server-db` if you are using the embedded database, and additional packages for plugins, depending on what was previously installed on the Server host. If the commands to update the server complete without errors, you can assume the upgrade has completed as desired. For additional assurance, you will have the option to check that the server versions have been updated after you start the server. The process of checking the server version is described in Step 5. Verify the Upgrade Succeeded.

Step 3. Start the Server

To start the server

On the Cloudera Manager Server host (the system on which you installed the `cloudera-manager-server` package) do the following:

If you are using the embedded PostgreSQL database for Cloudera Manager:

```
$ sudo service cloudera-scm-server-db start
```

This will set up the new database for Cloudera Navigator.

Note:

The `sudo service cloudera-scm-server-db start` command is not necessary if you are not using the embedded PostgreSQL database.

```
$ sudo service cloudera-scm-server start
```

You should see the following:

```
Starting cloudera-scm-server:                              [  OK  ]
```

Note:

If you have problems starting the server, such as database permissions problems, you can use the server's log `/var/log/cloudera-scm-server/cloudera-scm-server.log` to troubleshoot the problem.

Step 4. Deploy the Upgraded Software

Cloudera Manager can automatically upgrade existing agents. After you upgrade Cloudera Manager, when it is started for the first time, it checks for any older versions of agents. If older agents are detected, Cloudera Manager provides the opportunity to automatically update agents, which is recommended.

To upgrade the agents

Important:

All hosts in the cluster must have access to the Internet if you plan to use `archive.cloudera.com` as the source for installation files. If you do not have Internet access, create a custom repository.

1. Log in to the Cloudera Manager Admin Console. If you have just restarted the Cloudera Manager server, you may need to log in again.
2. On the Welcome screen, select whether you want to:
   - Install Cloudera Standard,
   - Try the Cloudera Enterprise with a 60-day trial license, or
   - Install a license you have purchased for Cloudera Enterprise.
3. After you upload the Cloudera Manager license, or if you have elected to use a Trial license, restart the Cloudera Manager server.

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

   - As the Cloudera Manager server restarts, the UI indicates its progress, and presents the login page when the restart has completed.
4. Click **Continue** to proceed to the Upgrade cluster hosts screen.
5. On the Upgrade cluster hosts screen, click **Start Upgrade** to upgrade the existing managed hosts. Click **Skip Host Upgrades** to skip this step.
6. Select the release of the Cloudera Manager Agent to install. Normally, this will be the **Matched Release for this Cloudera Manager Server**. However, if you used a custom repository for the Cloudera Manager server, select **Custom Repository** and provide the required information

   Click **Continue** to proceed.
7. Provide credentials for authenticating with hosts.
   a. Select **root** or enter the user name for an account that has password-less sudo permissions.
   b. Select an authentication method.
      - If you choose to use password authentication, enter and confirm the password.
      - If you choose to use public-key authentication provide a passphrase and path to the required key files.
      - You can choose to specify an alternate SSH port. The default value is 22.
      - You can specify the maximum number of host installations to run at once. The default value is 10.
8. Click **Start Installation** to install and start Cloudera Manager Agents. The status of installation on each host is displayed on the page that appears after you click **Start Installation**. You can also click the **Details** link for

individual hosts to view detailed information about the installation and error messages if installation fails on any hosts.

> **Note:**
>
> If you click the **Abort Installation** button while installation is in progress, it will halt any pending or in-progress installations and roll back any in-progress installations to a clean state. The **Abort Installation** button does not affect host installations that have already completed successfully or already failed.

If installation fails on a host, you can click the **Retry** link next to the failed host to try installation on that host again. To retry installation on all failed hosts, click **Retry Failed Hosts** at the bottom of the screen.

9. When the **Continue** button appears at the bottom of the screen, the installation process is complete. If the installation has completed successfully on some hosts but failed on others, you can click **Continue** if you want to skip installation on the failed hosts and continue to the next screen to start installing the Cloudera Management services on the successful hosts.
10. The Host Inspector runs to inspect your managed hosts for correct versions and configurations. If there are problems, you can make changes and them re-run the inspector. When you are satisfied with the inspection results, click **Continue** to install the Cloudera Management services.
11. On the next page, select the hosts where the Hive Metastore Server role should be installed.

    If you are upgrading from a version of Cloudera Manager prior to 4.5 this step will be skipped -- the Hive Metastore will already be set up.

    The Hive service is now managed by Cloudera Manager; you must select the host for the Hive Metastore Server. You should assign the Hive Metastore server to a single host.
12. Review the configuration values for your Hive roles, and click **Accept** to continue.

> **Note:**
>
> If Hue is using a Hive metastore of type Derby (the default), then the newly created Hive service will also use Derby. However, since Derby does not allow concurrent connection, the new Hive Metastore Server will fail to start. The failure is harmless the Hive Metastore Server is not used at this point) and intentional, to preserve the cluster functionality that existed before the upgrade.
>
> If you are upgrading to CM 4.5 or later from a release prior to 4.5 (i.e. 4.1 or earlier) Hive's metastore bypass mode is enabled by default. You should plan to disable the Bypass Hive Metastore Server mode, especially when using CDH 4.2 or later. Using the Hive Metastore Server is the recommended configuration. After changing this configuration, you must re-deploy your client configurations, restart Hive, and restart any Hue or Impala services configured to use that Hive.

13. Your services (except for Hive and the services you stopped in Step 1) should now be running.

Step 5. Verify the Upgrade Succeeded

If the commands to update and start the server complete without errors, you can assume the upgrade has completed as desired. For additional assurance, you can check that the server versions have been updated.

To verify the server upgrade succeeded

1. In the Cloudera Manager Admin console, click the **Hosts** tab.
2. Click **Host Inspector**. On large clusters, the host inspector may take some time to finish running. You must wait for the process to complete before proceeding to the next step.
3. Click **Show Inspector Results**.

All results from the host inspector process are displayed including the currently installed versions. If this includes listings of current component versions, the installation completed as expected.

Step 6. Add Hive Gateway Roles

You must add Hive Gateway roles to any hosts where Hive clients should run.

> **Note:**
>
> This step only applies if you are upgrading from a release prior to Cloudera Manager 4.5. If you are upgrading from 4.5 or later and you have Hive gateway roles already installed, you will not need to add them again.

To add Hive gateway roles:

1. In the Cloudera Manager Admin console, pull down the **Services** tab and select the Hive service.
2. Go to the **Instances** tab, and click the **Add** button. This opens the **Add Role Instances** page.
3. Select the hosts on which you want a Hive Gateway role to run. This will ensure that the Hive client configurations are deployed on these hosts.

Step 7. Restart Services

You must restart the Management Service and any other services (Hive, Hue, Impala) that you stopped at the beginning of this procedure. You should also restart the MapReduce service, or certain functions on MR roles will fail.

In addition, as of Cloudera Manager 4.1, health checks were introduced for the ZooKeeper service. If you are upgrading from a Cloudera Manager version older than 4.1 and have ZooKeeper installed, those new health checks will fail until you restart the ZooKeeper service.

To restart the ZooKeeper Service

1. From the **Services** tab select **All Services** in the Cloudera Manager Admin Console.
2. Choose **Restart** on the **Actions** menu for the ZooKeeper Service.

> **Note:**
>
> If for some reason you do not want to restart the ZooKeeper service at this point, you can disable the alerts for the failing health checks, or disable the health checks themselves. See Configuring Monitoring Settings. However, be sure to re-enable any checks you have disabled when you eventually restart the service. It is strongly recommended that you restart the service as soon as possible.

To start the services you stopped in Step 1:

1. From the **Services** tab select **All Services** in the Cloudera Manager Admin Console.
2. Choose **Start** on the **Actions** menu for the each service you need to start.

To start the Cloudera Management Service:

1. From the **Services** tab select **All Services** in the Cloudera Manager Admin Console.
2. Choose **Start** on the **Actions** menu for the Cloudera Management Services.

> **Note:**
>
> If you change the hostname or port where the Cloudera Manager is running, or you enable TLS security, you must restart the Cloudera Management Services to update the URL to the Server.

To restart the MapReduce Service:

1. From the **Services** tab in the Cloudera Manager Admin Console, select the MapReduce service.
2. Choose **Restart** on the **Actions** menu for the each service you need to start.

   If you do not restart MapReduce after an upgrade to Cloudera Manager 4.6,certain functions such as rolling restart, decommissioning TaskTrackers, or refreshing the JobTracker will fail. Once MapReduce has been restarted, these functions will work correctly from then on.

Test the Installation

When you have finished the upgrade to Cloudera Manager, you can test the installation to verify that the monitoring features are working as expected; follow instructions under Testing the Installation.

Adding the Cloudera Navigator Role

If you have upgraded to Cloudera Enterprise or are running the 60-day Trial and want to try Cloudera Navigator, you must add it as a role under the management service.

1. From the Services page, select the Management Service.
2. Go to the **Instances** tab, and click the **Add** button.
3. In the table presented, scroll to the end and select the host where you want the Navigator Server role to be hosted, and click **Continue**.
4. Because Cloudera Navigator is separately licensed, you are presented with a license statement. Click **Accept** to enable the trial license for this feature.
5. Enter the credentials for the database to be used by the Navigator Server. Assuming you have not set up an external database, you can use the Embedded Database for this. Click **Test Connection** to verify connectivity to the Database, the click **Continue**.
6. Review and accept any configuration changes (typically there are none). Click **Accept**. This returns you to the Instances page.
7. The Navigator Server role is added but not started. To start the role:

   a. Click the checkbox next to the role.
   b. From the **Actions for Selected** menu, click **Start**, and confirm that you want to start the role.

Step 8. Deploy Updated Client Configurations

During upgrades between major versions, resource locations may change. To ensure clients have current information about resources, update client configuration as described in Deploying Client Configuration Files.

Step 9. (Optional) Upgrade CDH

Cloudera Manager 4.x can manage both CDH3 and CDH4, so upgrading existing CDH3 installations is not required, but to get the benefits of CDH4, you may want to upgrade to the latest version. See the following topics for more information on upgrading CDH:

- Upgrading to the Latest Version of CDH4 in a Cloudera Managed Deployment - Follow this path to upgrade existing installations of CDH4 to the latest version of CDH4.
- Upgrading CDH3 to CDH4 in a Cloudera Managed Deployment - Follow this path to upgrade existing installations of CDH3 to the latest version of CDH4. You can also install Impala when you upgrade to CDH4 version 4.1.2 or later.

Upgrade from Cloudera Standard to Cloudera Enterprise

You have two options for upgrading from Cloudera Standard: you can upgrade temporarily using the 60-day Trial license, or you can upgrade permanently by installing a Cloudera Manager Enterprise license.

> **Note:** Cloudera Manager can continue to use the databases that were established for Cloudera Manager Standard. Unless you want to change databases, no database installations are required.

Upgrading your Cloudera Manager License

To upgrade your Cloudera Manager license:

1. From the **Administration**tab, select **License**.
2. Follow the instructions at Managing Licenses on page 51 to either upgrade to a full Enterprise license, or to start a 60-day Trial.

   > **Note:** You can only use the 60-day Trial option once.

Managing Licenses

When you install Cloudera Manager, you can choose to install Cloudera Standard (no license required), Cloudera Enterprise (which requires a license) or a 60-day trial of Cloudera Enterprise.

To access the **License** page, pull down the **Administration** menu and click **License** .

If you have a license installed, the license page indicates its status (for example, whether your license is currently valid) and shows you the owner, the license key, and the expiration date of the license, if there is one. This does not appear if you are running Cloudera Standard. Note that currently no Add-Ons are shown, even if you have licenses for Add-ons such as BDR, Navigator, RTD, or RTQ.

At the right side of the page a table shows the usage of licensed products based on the number of nodes with those products installed. Each cell in the table shows the number of copies of the product installed per cluster, as well as the total copies installed on nodes under management by this Cloudera Manager server. You move the cursor over the ? to see an explanation of each item.

- An **Enterprise Core node** is any host running one or more of the following: HDFS NameNode, HDFS DataNode, ZooKeeper Server, any non-Gateway Hive role.
- An **Enterprise BDR node** is any billable host in a cluster used for backup.
- An **Enterprise RTD node** is any host running one or more of the following: HBase RegionServer, HBase Master, ZooKeeper Server, HDFS NameNode.
- An **Enterprise RTQ node** is any host running one or more of the following: Impala StateStore, Impala Daemon, HDFS NameNode.
- A **Navigator node** is any billable host in a cluster managed by a CM instance running Navigator.

To End a 60-day Trial of Cloudera Enterprise

If you are using the **Trial Edition** the Details block indicates when your license will expire. However, you can end the trial at any time (prior to expiration) by clicking the **End Trial** button. The Enterprise-only features will be disabled the next time you log into Cloudera Manager.

To end a 60-day trial prior to its expiration date:

1. On the License page, click **End Trial**.
2. Confirm that you want to end the trial.
3. You must restart the Cloudera Manager server. Log in to your Cloudera Manager server host and restart the server from the command line.
4. When the server has restarted, the Cloudera Manager login page appears.
5. After you log back in, you may notice that the Reports Manager and Cloudera Navigator roles still appear. However, these roles are stopped, and cannot be restarted because they are unlicensed.

> **Note:** When your 60-day trial ends, features will continue to work until you restart the Cloudera Manager server. However, you will not be able to log in again until you restart the server. However, data or configurations associated with the disabled functions will not be deleted, and will become available again if you install an Enterprise license. Trial expiration (or termination) will have the following effects:
>
> - Only local users will be able to log in (no LDAP authentication).
> - Configuration History will be unavailable.
> - SNMP alerts will not longer occur.
> - Operation Reports will be inaccessible (but will remain in the database).
> - Replication jobs (available with BDR) will no longer run.
> - Commands such as Rolling Restart, History and Rollback (under the Configuration tab), Send Diagnostic Data, Replication, and starting the Navigator role will not be available or will be disabled.

To Upgrade from Cloudera Standard to a 60-day Trial of Cloudera Enterprise

> **Note:** You can only use the 60-day Trial option only once. Once the trial period has expired, or you have ended the trial, you cannot restart it.

If you are using Cloudera Manager without the Enterprise license, you will not have access to certain features that are part of Cloudera Enterprise. If you have not previously used the Free Trial option, you can start a 60-day trial from the License page.

To start a 60-day Trial:

1. On the License page, click **Try Cloudera Enterprise for 60 Days**.
2. Cloudera Manager presents a pop-up describing the features enabled with Cloudera Enterprise. Click **OK** to proceed.
3. You must restart the Cloudera Manager server for your trial license to take effect. Log in to your Cloudera Manager server host and restart the server from the command line.
4. When the server has restarted, the Cloudera Manager login page appears.
5. After you log in, you must configure some of the additional roles that are enabled.
   a. Designate a host for the Management Service Reports Manager role.
   b. Enter the credentials for the database to be used by the Reports Manager. Assuming you have not set up an external database, you can use the Embedded Database for this. Click **Test Connection** to verify connectivity to the Database, the click **Continue**.
   c. Review and accept any configuration changes (typically there are none).
6. At this point, your installation is upgraded. However, the Cloudera Navigator role, which is a separately-licensed product, is not automatically added. If you have a license for Cloudera Navigator, you must add its role to the set of Cloudera Manager management services. See the instructions in the next section (Adding the Cloudera Navigator Role).
7. You may need to restart services if they have outdated configurations. It is also recommended that you redeploy your client configuration files.

Adding the Cloudera Navigator Role

If you want to try Cloudera Navigator, you must add it as a role under the management service.

1. From the Services page, select the Management Service.
2. Go to the **Instances** tab, and click the **Add** button.
3. In the table presented, scroll to the end and select the host where you want the Navigator Server role to be hosted, and click **Continue**.

4. Because Cloudera Navigator is separately licensed, you are presented with a license statement. Click **Accept** to enable the trial license for this feature.
5. Enter the credentials for the database to be used by the Navigator Server. Assuming you have not set up an external database, you can use the Embedded Database for this. Click **Test Connection** to verify connectivity to the Database, the click **Continue**.
6. Review and accept any configuration changes (typically there are none). Click **Accept**. This returns you to the Instances page.
7. The Navigator Server role is added but not started. To start the role:

   a. Click the checkbox next to the role.
   b. From the **Actions for Selected** menu, click **Start**, and confirm that you want to start the role.

To upgrade from Cloudera Standard to Cloudera Enterprise

You can upgrade to Cloudera Enterprise by uploading a license key purchased from Cloudera.

1. From the License page, click **Upload a Cloudera Enterprise License**.
2. Cloudera Manager presents a pop-up the presents the caveats concerning the separately-licensed products. Click **Upload License** to proceed.
3. You must restart the Cloudera Manager server for your Enterprise license to take effect. Log in to the Cloudera Manager server host and restart the server from the command line.
4. When the server has restarted, the Cloudera Manager login page appears.
5. After you log in, you must configure some of the additional roles that are enabled.

   a. Designate a host for the Management Service Reports Manager role.
   b. Enter the credentials for the database to be used by the Reports Manager. Assuming you have not set up an external database, you can use the Embedded Database for this. Click **Test Connection** to verify connectivity to the Database, the click **Continue**.
   c. Review and accept any configuration changes (typically there are none).

6. At this point, your installation is upgraded. However, Cloudera Navigator, which is a separately-licensed product, is not automatically added. See the section above (Adding the Cloudera Navigator Role) for instructions.
7. You may need to restart services if they have outdated configurations. It is also recommended that you redeploy your client configuration files.

Configuring Alert Delivery

Under the Alert Publisher role of the Cloudera Manager Management Service, you can configure email or SNMP delivery of alert notifications.

Configuring Alert Email Delivery

When you install the Cloudera Manager Management Services, it asks you for information about the mail server you will use with the Alert Publisher. However, if you need to change these settings, you can do so under the Alert Publisher section of the Management Services configuration tab.

Note that if you just want to add to or modify the list of alert recipient email addresses, you can do from the **Alerts** page, accessed under the **Administration** tab.

You can also send a test alert e-mail from the **Alerts** page under the **Administration** tab.

You can enable and disable email alerts delivery entirely (without changing the other email settings) with the **Enable email alerts** property.

To enable, disable, or configure email alerts:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.
2. Select **Configuration** > **View and Edit** .
3. Select the **Alert Publisher (Default)** role group to see the list of properties. In order to receive email alerts you must set (or verify) the following settings:
    - Email protocol to use.
    - Your mail server hostname and port.
    - The username and password of the email user that will be logged into the mail server as the "sender" of the alert emails.
    - A comma-separated list of email addresses that will be the recipients of alert emails.
    - The format of the email alert message. Select **json** if you need the message to be parsed by a script or program.
4. Click the **Save Changes** button at the top of the page to save your settings.
5. You will need to restart the Alert Publisher role to have these changes take effect.

The following pages have more details on configuring SNMP and alerts:

- Configuring SNMP
- Alert Settings

Configuring SNMP

> **Note: This feature is available only with Cloudera Enterprise.**
>
> The feature described in this section is not available in Cloudera Manager with Cloudera Standard.
>
> If you have been using the Cloudera Enterprise Trial Edition, this feature will no longer be available after your trial license expires.
>
> To obtain a license for Cloudera Enterprise, please contact sales@cloudera.com. When you install your Enterprise license, this feature will be enabled.

Before you enable SNMP traps, make sure you have configured your trap receiver (Network Management System or SNMP server) with the Cloudera MIB.

To view the Cloudera MIB:

1. From the All Services page, go to the Cloudera Manager management service.
2. From the **Configuration** tab select **View and Edit**.
3. Expand the **Alert Publisher** category in the Category list at the left, and select **SNMP**.
4. In the **Description** column for the first property (**SNMP NMS Hostname**) there is a link to the **SMNP MIB**.
5. Click the link in the Description field to view the MIB.

To enable, disable, or configure SNMP traps:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.
2. Pull down the **Configuration** tab and click **Edit**.
3. Under the **Alert Publisher (Base)** role group select **SNMP** to see the list of properties.
   - Enter the DNS name or IP address of the Network Management System (SNMP server) acting as the trap receiver in the SNMP NMS Hostname property.
   - Select the version of SNMP you are using: SNMPv2, SNMPv3 authentication with no privacy (`authNoPriv`), or SNMPv3 with no authentication and no privacy (`noAuthNoPriv`).
   - For SNMPv2, you must enter a Community String.
   - For SNMPv3, you must enter the SNMP Server Engine ID.
   - For SNMPv3 with authentication (`authNoPriv`) you must also enter the Security user name, Authentication protocol, and protocol pass phrase.
   - You can also change other settings such as the port, retry, or timeout values.
4. Click **Save Changes** when you are done.
5. You must restart the Alert Publisher role to have these changes take effect.

To disable SNMP traps, simply remove the hostname from the SNMP NMS Hostname property (alert.snmp.server.hostname).

Alert Settings

The **Alerts** page (found under the Administration tab) provides a summary of the settings for alerts in your clusters.

- Pull down the **Administration** tab and select **Alerts**.

Alert Type The left column lets you select by alert type (Health, Log, or Activity) and within that by service instance. In the case of Health alerts, you can look at alerts for Hosts as well. You can select an individual service to see just the alert settings for that service.

Health/Log/Activity Alert Settings Depending on your selection in the left column, the right hand column show you the list of alerts that are enabled or disabled for the selected service type.

To change the alert settings for a service, click the edit icon (✎) next to the service name. This will take you to the Monitoring section of the Configuration tab for the service. From here you can enable or disable alerts and configure thresholds as needed.

Recipients You can also view the list of recipients configured for the enabled alerts. Again, click the edit icon (✎) at the top of this list to go to the Alert Publisher configuration settings, where you can modify the list of recipients.

If you want to verify that the recipients will actually receive an alert, click the **Send Test Alert** link under the list of recipients. This will send a test alert to all recipients in the list.

Configuring Management Services Database Limits

Each Cloudera Management Service 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. For example, the Activity Monitor maintains data at the service level, the activity level (MapReduce jobs and aggregate activities), and at the task attempt level. Limits on these data sets are configured when you install your management services, but you can modify these parameters through the Configuration settings in the Cloudera Manager Admin console, for each management service.

For example, the Event Server lets you set a total number of events you want to store. Host Monitor and Service Monitor let you set data expiration thresholds (in hours), and Activity Monitor gives you "purge" settings (also in hours) for the data it stores. 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.

To change any of the data retention or log size settings:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.
2. Select **Configuration** > **View and Edit** .
3. In the left-hand column, select the role group for the role whose configurations you want to modify. (Note that the management services are singleton roles so there will be only a Base role group for the role.)
4. For some services, such as the Activity Monitor, Service Monitor, or Host Monitor, the purge or expiration period properties are found in the top-level settings for the role. Typically, Log file size settings will be under the **Logs** category under the role group.

Other Cloudera Manager Settings

From the **Administration** tab you can select options for configuring settings that affect how Cloudera Manager interacts with your cluster.

The Administration Settings Page

The **Settings** page provides a number of categories as follows:

- **Performance** — Set the Cloudera Manager Agent heartbeat here.
- **Advanced** — Enable API debugging and other advanced options.
- **Thresholds** — Set Agent Health status parameters. For configuration instructions, see Configuring Agent Heartbeat and Health Status Options.
- **Security** — Set TLS encryption settings to enable TLS encryption between the Cloudera Manager Server, Agents, and clients. For configuration instructions, see Configuring TLS Security for Cloudera Manager You can also:
  - Set the realm for Kerberos security and point to a custom keytab retrieval script. For configuration instructions, see Configuring Hadoop Security with Cloudera Manager.
  - 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 the Ports for the Admin Console and Agents.
- **Other** — To enable Cloudera usage data collection For configuration instructions, see Configuring Anonymous Usage Data Collection on page 65. You can also:
  - 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. The user must click "I Agree" in order to proceed to the login dialog.
  - Disable/enable the auto-search for the Events panel at the bottom of a page.
- **Support** — Enable access to online Help files from the Cloudera web site rather than from locally-installed files. (see Opening the Help Files from the Cloudera Web Site on page 61), and enable automatic sending of diagnostic data to Cloudera when you trigger a data collection (see Sending Diagnostic Data to Cloudera)
- **External Authentication** — Specify the configuration to use LDAP, Active Directory, or an external program for authentication. See Configuring External Authentication for instructions.
- **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 Parcel Configuration Settings for more information.

Opening the Help Files from the Cloudera Web Site

By default, when you click the Help link under the Support menu in the Cloudera Manager Admin console, Help files from the Cloudera web site are opened. This is because local Help files are not updated after installation. You can configure Cloudera Manager to open either the latest Help files from the Cloudera web site (this option requires Internet access from the browser) or locally-installed Help files.

To configure Cloudera Manager to open the Help files from the Cloudera web site (or local Help files):

1. From the **Administration** tab, select **Settings**.
2. Under the **Support** category, enable the **Open latest Help files from the Cloudera website**. This setting will be enabled by default and you can uncheck this option to open the locally-installed Help documents.
3. Click **Save Changes**.

User Interface Language Settings

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 > Language page. You can also change the language for the information provided with activity and health events, and for alert email messages.

To change the language of the activity and health event information and alert email messages, select the language you want from the drop-down list on this page, then click **Save Changes**.

Kerberos

After enabling and configuring Hadoop security using Kerberos on your cluster, you can view and regenerate the Kerberos principals for your cluster. If you make a global configuration change in your cluster, such as changing the encryption type, you would use the Kerberos page to regenerate the principals for your cluster.

In a secure cluster, the Kerberos page lists all the Kerberos principals that are active on your cluster.

Regenerating your Kerberos Principals

If you make a global configuration change in your cluster, such as changing the encryption type, you must use the following instructions to regenerate the principals for your cluster.

Important:

Do not regenerate the principals for your cluster unless you have made a global configuration change. Before regenerating, be sure to read the Set up a Local KDC and Default Domain for the Hadoop Cluster section to avoid making your existing host keytabs invalid.

To view and regenerate the Kerberos principals for your cluster:

1. From the **Administration** tab, select **Kerberos**.
2. The currently configured Kerberos principals are displayed. If you are running HDFS, the `hdfs/hostname` and `host/hostname` principals are listed. If you are running MapReduce, the `mapred/hostname` and `host/hostname` principals are listed. The principals for other running services are also listed.
3. Only if necessary, select the principals you want to regenerate.
4. Click **Regenerate**.

The Security Inspector

The Security Inspector uses the Host Inspector to run a security-related set of commands on the hosts in your cluster. It reports on things such as how java is configured for encryption, and reports on the default realms configured on each host.

To use the Security Inspector:

1. Under the Administration tab, select **Kerberos**.
2. Click **Security Inspector**. Cloudera Manager begins several tasks to inspect the managed hosts.
3. After the inspection completes, click **Download Result Data** or **Show Inspector Results** to review the results.

Importing Cloudera Manager Settings

Backing up your Current Deployment

To back up your current deployment, please see the section on backing up your database in Database Considerations for Cloudera Manager Upgrades on page 30. The import feature should not be relied on for backup and recovery at this time.

Building a Cloudera Manager Deployment

You can use the Cloudera Manager API to programmatically build a Cloudera Manager Deployment — a definition of all the entities in your Cloudera Manager-managed deployment — clusters, service, roles, hosts, users and so on. See the Cloudera Manager API documentation on how to manage deployments using the `/cm/deployment` resource.

Uploading a Cloudera Manager 4.0 Configuration Script

> **Note:** As of Cloudera Manager 4.1, the import of configuration settings through the Cloudera Manager Admin Console UI has been deprecated. If you have exported a configuration using the **Export** tab in an older version of Cloudera Manager, you can still import it following the instructions below. However, going forward, importing a deployment should be done using the Cloudera Manager API. See the documentation for /cm/deployment for details.

> **Important:** You must import the configuration settings on a clean cluster that does not have existing hosts or services.

> **Important:** When you first installed the Cloudera Manager Server, you set up a database to store the Cloudera Manager service configuration information (see Installing and Configuring Databases). That database also stores the Cloudera Manager license information. If the original database is lost (for example, the database was deleted and you recreated a new one), you must first upload your license on the **Administration > License** page and restart the Cloudera Manager Server before importing the configuration settings. If you don't upload your license first to store the license information in the new database, the import will fail.

To import the configuration script into Cloudera Manager:

1. On every Cloudera Manager Agent host, run this command to stop the Cloudera Manager Agent:

```
$ sudo service cloudera-scm-agent stop
```

2. Delete all services on the **Services** tab by choosing **Delete** from the **Actions** menu next to each service instance.
3. Delete all hosts on the **Hosts** tab by clicking the check box at the top of list of hosts, and then click **Delete**.
4. Copy the configuration script file that you downloaded during export to the host with the new Cloudera Manager server.

5. Pull down the **Administration** tab and select **Import**.
6. Click **Browse**, navigate to the file location, and click **Open**.
7. Click **Import**.
8. On every Cloudera Manager Agent host, run this command to start the Cloudera Manager Agent:

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

Sending Usage and Diagnostic Data to Cloudera

With Cloudera Manager you can take regularly-scheduled snapshots of the state of your cluster and automatically send it anonymously to Cloudera. This helps Cloudera improve and optimize Cloudera Manager. If you are a Cloudera Enterprise user, you can also trigger the collection of diagnostic data and send it to Cloudera Support to aid in resolving a problem you may be having. The following sections provide more information about these features.

- Configuring Anonymous Usage Data Collection
- Sending Diagnostic Data to Cloudera

Configuring Anonymous Usage Data Collection

You can configure Cloudera Manager to send anonymous usage information using Google Analytics to Cloudera. The information helps Cloudera improve Cloudera Manager.

To configure anonymous usage data collection:

1. From the **Administration** tab, select **Settings** .
2. Under the **Other** category, set the **Allow Usage Data Collection** option to enable or disable anonymous usage data collection.
3. Click **Save Changes**.

Sending Diagnostic Data to Cloudera

To help with solving problems when using Cloudera Manager on your cluster, you can collect diagnostic data on a regular schedule, and have it automatically sent to Cloudera. Cloudera Manager is configured by default to collect data weekly and to send it automatically. You can schedule the frequency of data collection on a daily, weekly, or monthly schedule, or disable the scheduled collection of data entirely. Separately you can disable the automatic sending of data to Cloudera — see Disabling the Automatic Sending of Diagnostic Data. You can also send a collected data set manually.

If you are a Cloudera Enterprise customer, you can also trigger a collection and send the resulting diagnostic data to Cloudera Support on demand to aid in diagnosing problems you may be having.

> **Note:**
>
> To automatically send diagnostic data requires the Cloudera Manager Server host to have Internet access, and be configured for sending data automatically. If your Cloudera Manager server does not have Internet access, you can manually send the diagnostic data as described below.

This section covers the following topics:

- Configuring the Frequency of Diagnostic Data Collection
- Collecting and Sending Diagnostic Data to Cloudera on Demand
- Disabling the Automatic Sending of Diagnostic Data
- Manually Sending Diagnostic Data to Cloudera
- What Data Does Cloudera Manager Collect?

Configuring the Frequency of Diagnostic Data Collection

By default, Cloudera Manager collects diagnostic data on a weekly basis, and automatically sends it to Cloudera. You can change the frequency to daily, weekly, monthly, or never. If you are a Cloudera Enterprise customer and you set the schedule to Never you can still collect and send data to Cloudera on demand.

To change the frequency of diagnostic data collection:

1. From the **Administration** tab, select **Settings**.
2. Under the **Support** category, click in the field for the property **Send diagnostic Data to Cloudera Automatically** and select the frequency you want.
3. You can change the day and time of day that the collection will be performed.
4. Click **Save Changes**

You can see the setting for the current data collection frequency under the **Support** menu in the main navigation bar.

Collecting and Sending Diagnostic Data to Cloudera on Demand

> **Note: This feature is available only with Cloudera Enterprise.**
>
> The feature described in this section is not available in Cloudera Manager with Cloudera Standard.
>
> If you have been using the Cloudera Enterprise Trial Edition, this feature will no longer be available after your trial license expires.
>
> To obtain a license for Cloudera Enterprise, please contact sales@cloudera.com. When you install your Enterprise license, this feature will be enabled.

As a Cloudera Enterprise customer, you can have Cloudera Manager collect a set of diagnostic If you do not want data sent automatically, you must disable that feature (see Disabling the Automatic Sending of Diagnostic Data).

To collect and send diagnostic data to Cloudera:

1. Click the **Support** menu link.
2. Choose **Send Diagnostic Data**. This opens the Send Diagnostic Data form. Note that at the top of the form it tells you whether Cloudera Manager is configured to send the data automatically or not. See the instructions below to change this.
3. Fill in or change the information here as appropriate.
   - To change the System Identifier, from the **Administration** tab select **Settings** and go to the**Other** category.
   - Cloudera Manager pre-populates the End Time based on the setting of the Time Range Selector. You should change this to be a few minutes after you observed the problem or condition that you are trying to capture. Note that the time range is based on the time zone of the host where Cloudera Manager server is running.
   - If you have a support ticket open with Cloudera support, please include the support ticket number in the field provided.
4. Click **Collect Diagnostic Data**. A Running Commands window shows you the progress of the data collection steps. When these steps are complete, the collected data is sent to Cloudera.

Disabling the Automatic Sending of Diagnostic Data

If you do not want data sent to Cloudera automatically, you can disable this feature. The data you collect will be saved

To disable sending diagnostic data to Cloudera automatically:

1. From the **Administration** tab, select **Settings**.
2. Under the **Support** category, uncheck the box for **Send diagnostic Data to Cloudera Automatically**.
3. Click **Save Changes**

Manually Sending Diagnostic Data to Cloudera

> **Note: This feature is available only with Cloudera Enterprise.**
>
> The feature described in this section is not available in Cloudera Manager with Cloudera Standard.
>
> If you have been using the Cloudera Enterprise Trial Edition, this feature will no longer be available after your trial license expires.
>
> To obtain a license for Cloudera Enterprise, please contact sales@cloudera.com. When you install your Enterprise license, this feature will be enabled.

> **Note:**
>
> Automatically sending diagnostic data may fail sometimes and return an error message of "Could not send data to Cloudera." To work around this issue, you can manually send the data to Cloudera Support as described below.

To manually send collected diagnostic data to Cloudera:

1. Click the **Support** menu link.
2. Choose **Send Diagnostic Data**. This opens the Send Diagnostic Data form. Note that at the top of the form it tells you whether Cloudera Manager is configured to send the data automatically or not. See the instructions above to change this.
3. Fill in or change the information in the form as appropriate. Cloudera Manager pre-populates the start and end times, but you can change them. If you have a support ticket open with Cloudera support, please include the support ticket number in the field provided.
4. Click **Collect Diagnostic Data**. A Command Details window shows you the progress of the data collection steps.
5. Click **Download Result Data** to download and save a zip file of the information collected, on a host that has Internet access.
6. Download this script and run the following command on that host to send the data to Cloudera Support:

```
python phone_home.py --file [file-you-downloaded]
```

> **Note:**
>
> If you want to send your file manually but choose not to download the script, you can follow the instructions documented on the Cloudera Customer Portal at Get Support - Uploading Files for Cloudera Support.

What Data Does Cloudera Manager Collect?

Cloudera Manager collects and returns a significant amount of information about the health and performance of the cluster. It includes the following:

- Up to 1000 Cloudera Manager Audit Events: Configuration changes, add/remove of users, roles, services, etc.

- Data about the cluster structure which includes a list of all hosts, roles, and services along with the configs that are set through Cloudera Manager. Where passwords are set in Cloudera Manager, the passwords are not returned.
- Cloudera Manager License and version number.
- One day's worth of Cloudera Manager events: This includes critical errors Cloudera Manager watches for and more.
- Current health information for hosts, service, and roles. Includes results of health tests run by Cloudera Manager.
- Heartbeat information from each host, service, and role. These include status and some information about memory/disk/processer usage.
- The results of running Host Inspector.
- One day's worth of Cloudera Manager metrics.

> **Note:** If you are using Cloudera Standard, Host metrics are not included.

- A download of the debug pages for Cloudera Manager roles.
- For each machine in the cluster, the result of running a number of system-level commands on that machine.
- Logs from each role on the cluster, as well as the CM server and agent logs.