

Cloudera Upgrade



Important Notice

© 2010-2017 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, 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. For information about patents covering Cloudera products, see <http://tiny.cloudera.com/patents>.

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, Bldg 3

Palo Alto, CA 94304

info@cloudera.com

US: 1-888-789-1488

Intl: 1-650-362-0488

www.cloudera.com

Release Information

Version: Cloudera Enterprise 5.11.x

Date: May 10, 2017

Table of Contents

Cloudera Upgrade Overview.....	8
Assessing the Impact of an Upgrade.....	8
Overview of Upgrading Cloudera Manager.....	9
Overview of Upgrading CDH.....	10
Overview of Upgrading Cloudera Navigator Components.....	10
Upgrading the JDK.....	11
Upgrading Cloudera Manager.....	12
Upgrading Cloudera Manager 5 Using Packages.....	12
<i>Step 1: Collect Upgrade Information.....</i>	<i>13</i>
<i>Step 2: Complete Pre-Upgrade Steps.....</i>	<i>14</i>
<i>Step 3: Back Up Cloudera Manager Databases.....</i>	<i>14</i>
<i>Step 4: Upgrade the JDK</i>	<i>15</i>
<i>Step 5: Establish Access to the Software.....</i>	<i>15</i>
<i>Step 6: Prepare the Cloudera Navigator Data Management Component for Upgrade.....</i>	<i>15</i>
<i>Step 7: Upgrade the Cloudera Manager Server.....</i>	<i>16</i>
<i>Step 8: Verify and Test the Upgrade.....</i>	<i>24</i>
<i>Step 9: Upgrade any Required Navigator Components.....</i>	<i>25</i>
<i>Step 10: (Optional) Upgrade CDH.....</i>	<i>25</i>
Upgrading Cloudera Manager 5 Using Tarballs.....	25
<i>Step 1: Collect Upgrade Information.....</i>	<i>26</i>
<i>Step 2: Complete Pre-Upgrade Steps.....</i>	<i>26</i>
<i>Step 3: Upgrade the JDK.....</i>	<i>27</i>
<i>Step 4: Creating and Using a Package Repository for Cloudera Manager.....</i>	<i>27</i>
<i>Step 5: Upgrading the Cloudera Navigator Data Management Component.....</i>	<i>27</i>
<i>Step 6: Upgrading the Cloudera Manager Server and Agents.....</i>	<i>27</i>
<i>Step 7: Verify and Test the Upgrade.....</i>	<i>31</i>
<i>Step 8: Upgrading Cloudera Navigator Components.....</i>	<i>32</i>
<i>Step 9: (Optional) Upgrade CDH.....</i>	<i>32</i>
Upgrading Cloudera Manager 4 to Cloudera Manager 5.....	32
Review Warnings and Notes.....	33
Before You Begin.....	34
Stop Selected Services and Roles.....	35
Stop Cloudera Manager Server, Database, and Agent.....	35
(Optional) Upgrade the JDK on Cloudera Manager Server and Agent Hosts.....	36
Upgrade Cloudera Manager Software.....	36
Start Cloudera Manager Server.....	41

<i>Upgrade and Start Cloudera Manager Agents.....</i>	<i>42</i>
<i>Verify that the Upgrade Succeeded.....</i>	<i>48</i>
<i>Add Hive Gateway Roles.....</i>	<i>48</i>
<i>Configure the Cluster Version for Package Installs.....</i>	<i>48</i>
<i>Upgrade Impala.....</i>	<i>48</i>
<i>(Optional) Configure TLS/SSL for Cloudera Management Service.....</i>	<i>48</i>
<i>(Optional) Deploy a Cloudera Manager Agent Upgrade.....</i>	<i>49</i>
<i>Deploy a JDK Upgrade.....</i>	<i>50</i>
<i>(Optional) Deploy a Monitoring Upgrade.....</i>	<i>51</i>
<i>Start Selected Services and Roles.....</i>	<i>51</i>
<i>Deploy Updated Client Configurations.....</i>	<i>51</i>
<i>Test the Installation.....</i>	<i>51</i>
<i>(Optional) Upgrade CDH.....</i>	<i>52</i>
<i>Package Dependencies.....</i>	<i>52</i>

Upgrading CDH and Managed Services Using Cloudera Manager.....54

<i>Upgrading CDH 5.x.....</i>	<i>54</i>
<i>Minor Version Upgrades.....</i>	<i>54</i>
<i>Maintenance Version Upgrades.....</i>	<i>54</i>
<i>Upgrading CDH 4 to CDH 5.....</i>	<i>55</i>
<i>Upgrading CDH 4.x.....</i>	<i>55</i>
<i>Upgrading CDH 3.....</i>	<i>55</i>
<i>Upgrading to CDH 5.x Using a Rolling Upgrade.....</i>	<i>55</i>
<i>Step 1: Collect Upgrade Information.....</i>	<i>56</i>
<i>Step 2: Complete Pre-Upgrade Steps.....</i>	<i>56</i>
<i>Step 3: Ensure High Availability Is Enabled.....</i>	<i>58</i>
<i>Step 4: Back Up HDFS Metadata.....</i>	<i>58</i>
<i>Step 5: Back Up Databases.....</i>	<i>58</i>
<i>Step 6: Run the Upgrade Wizard.....</i>	<i>59</i>
<i>Step 7: Recover from Failed Steps or Perform a Manual Upgrade.....</i>	<i>60</i>
<i>Step 8: Remove the Previous CDH Version Packages and Refresh Symlinks.....</i>	<i>60</i>
<i>Step 9: Finalize HDFS Rolling Upgrade.....</i>	<i>61</i>
<i>Step 10: Exit Maintenance Mode.....</i>	<i>61</i>
<i>Upgrading to CDH 5.x Using Parcels.....</i>	<i>62</i>
<i>Step 1: Collect Upgrade Information.....</i>	<i>62</i>
<i>Step 2: Complete Pre-Upgrade Steps.....</i>	<i>62</i>
<i>Step 3: Stop Cluster Services.....</i>	<i>64</i>
<i>Step 4: Back up the HDFS Metadata on the NameNode.....</i>	<i>64</i>
<i>Step 5: Back Up Databases.....</i>	<i>65</i>
<i>Step 6: Run the Upgrade Wizard.....</i>	<i>66</i>
<i>Step 7: Recover from Failed Steps or Perform a Manual Upgrade.....</i>	<i>67</i>
<i>Step 8: Remove the Previous CDH Version Packages and Refresh Symlinks.....</i>	<i>67</i>
<i>Step 9: Finalize the HDFS Metadata Upgrade.....</i>	<i>68</i>

<i>Step 10: Exit Maintenance Mode</i>	68
Upgrading to CDH 5.x Using Packages.....	68
<i>Step 1: Collect Upgrade Information</i>	69
<i>Step 2: Complete Pre-Upgrade Steps</i>	69
<i>Step 3: Upgrade Unmanaged Components</i>	70
<i>Step 4: Stop Cluster Services</i>	71
<i>Step 5: Back up the HDFS Metadata on the NameNode</i>	71
<i>Step 6: Back Up Databases</i>	71
<i>Step 7: Upgrade Managed Components</i>	72
<i>Step 8: Update Symlinks for the Newly Installed Components</i>	75
<i>Step 9: Run the Upgrade Wizard</i>	75
<i>Step 10: Recover from Failed Steps or Perform a Manual Upgrade</i>	76
<i>Step 11: Finalize the HDFS Metadata Upgrade</i>	76
<i>Step 12: Exit Maintenance Mode</i>	77
Upgrade Managed Components Using a Specific Set of Packages.....	77
Performing Upgrade Wizard Actions Manually.....	80
Upgrading to CDH 5.8.0 or CDH 5.8.1 When Using the Flume Kafka Client.....	81
Upgrading CDH 4 to CDH 5.....	83
Upgrading from CDH 4 to CDH 5 Parcels.....	87
Upgrading from CDH 4 Packages to CDH 5 Packages.....	95
Upgrading CDH 4.....	104
Upgrading CDH 4 Using Parcels.....	105
Upgrading CDH 4 Using Packages.....	108
Performing a Rolling Upgrade on a CDH 4 Cluster.....	113
Upgrading CDH 3.....	116
Rolling Back a CDH 4-to-CDH 5 Upgrade.....	117
Backing Up Before Upgrading from CDH 4 to CDH 5	117
Procedure for Rolling Back a CDH 4-to-CDH 5 Upgrade	121

Upgrading to Oracle JDK 1.7.....130

Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment.....	130
Upgrading to Oracle JDK 1.7 in an Unmanaged Deployment.....	131

Upgrading to Oracle JDK 1.8.....132

Upgrading to Oracle JDK 1.8 in a Cloudera Manager Deployment.....	132
Upgrading to Oracle JDK 1.8 in an Unmanaged Deployment.....	133
Using AES-256 Encryption.....	133

Upgrading Cloudera Navigator Components.....135

Upgrading the Cloudera Navigator Data Management Component.....	135
Upgrading Cloudera Navigator Key Trustee Server.....	136

<i>Upgrading Cloudera Navigator Key Trustee Server 3.x to 5.4.x.....</i>	<i>136</i>
<i>Upgrading Cloudera Navigator Key Trustee Server 3.8 to 5.5 Using the ktupgrade Script.....</i>	<i>141</i>
<i>Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher.....</i>	<i>146</i>
<i>Upgrading Cloudera Navigator Key HSM.....</i>	<i>151</i>
<i>Setting Up an Internal Repository.....</i>	<i>151</i>
<i>Upgrading Key HSM.....</i>	<i>151</i>
<i>Upgrading Key Trustee KMS.....</i>	<i>152</i>
<i>Setting Up an Internal Repository.....</i>	<i>152</i>
<i>Upgrading Key Trustee KMS Using Parcels.....</i>	<i>152</i>
<i>Upgrading Key Trustee KMS Using Packages.....</i>	<i>153</i>
<i>Upgrading Cloudera Navigator Encrypt.....</i>	<i>153</i>
<i>Setting Up an Internal Repository.....</i>	<i>153</i>
<i>Upgrading Navigator Encrypt (RHEL-Compatible).....</i>	<i>153</i>
<i>Upgrading Navigator Encrypt (SLES).....</i>	<i>154</i>
<i>Upgrading Navigator Encrypt (Debian or Ubuntu).....</i>	<i>155</i>
<i>Best Practices for Upgrading Navigator Encrypt Hosts.....</i>	<i>155</i>

Creating and Using a Package Repository.....157

<i>Creating a Permanent Remote Repository.....</i>	<i>157</i>
<i>Installing a Web Server.....</i>	<i>157</i>
<i>Creating a Temporary Remote Repository.....</i>	<i>158</i>
<i>Modifying Clients to Find the Repository.....</i>	<i>158</i>

Database Considerations for Cloudera Manager Upgrades.....160

<i>Backing up Databases.....</i>	<i>160</i>
<i>Creating New Databases.....</i>	<i>160</i>
<i>Modifying Databases to Support UTF-8.....</i>	<i>160</i>
<i>Modifying Databases to Support Appropriate Maximum Connections.....</i>	<i>161</i>
<i>Next Steps.....</i>	<i>162</i>

Re-Running the Cloudera Manager Upgrade Wizard.....163

Reverting a Failed Cloudera Manager Upgrade.....164

<i>Reinstall the Cloudera Manager Server Packages.....</i>	<i>164</i>
<i>Start the Server.....</i>	<i>166</i>

Upgrading Unmanaged CDH Using the Command Line.....167

<i>Upgrading from CDH 4 to CDH 5.....</i>	<i>167</i>
<i>Before You Begin Upgrading to CDH 5 Using the Command Line.....</i>	<i>167</i>
<i>Upgrading to CDH 5.....</i>	<i>169</i>

Upgrading from an Earlier CDH 5 Release to the Latest Release.....	184
<i>Important Tasks.....</i>	<i>184</i>
<i>Before Upgrading to the Latest Release of CDH.....</i>	<i>185</i>
<i>Upgrading from CDH 5.4.0 or Higher to the Latest Release.....</i>	<i>186</i>
<i>Upgrading from a Release Lower than CDH 5.4.0 to the Latest Release.....</i>	<i>199</i>

Upgrading Host Operating Systems in a CDH Cluster.....215

Cloudera Upgrade Overview

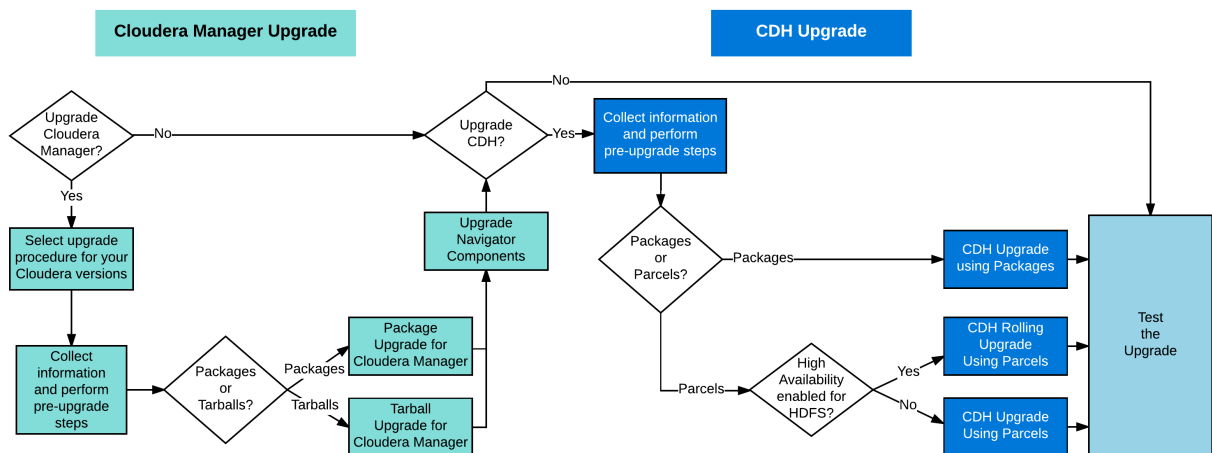
This topic provides an overview of upgrade procedures for Cloudera Manager and CDH.

The procedures described here are for clusters managed by Cloudera Manager. For information about upgrading unmanaged CDH clusters (clusters that are not managed by Cloudera Manager), see [Upgrading Unmanaged CDH Using the Command Line](#) on page 167.

When upgrading Cloudera Manager, you can use tarballs or operating system packages. When upgrading CDH, you can use packages or parcels. You might also need to install a new version of the JDK. Cloudera Navigator is also upgraded when you upgrade Cloudera Manager.

You are not required to upgrade Cloudera Manager and CDH at the same time, but the versions of Cloudera Manager and CDH must be compatible. Cloudera Manager can manage clusters for the current and previous major versions of CDH and any equal or lower minor version of CDH. For example, Cloudera Manager 5.7.1 can manage clusters with CDH 5.7.2, CDH 5.6.1, and CDH 4.8.6 but cannot manage a cluster with CDH 5.8.1. Cloudera Manager 5.x cannot manage clusters using CDH 3.x.

You can choose from several options as you upgrade:



Assessing the Impact of an Upgrade

Plan for a sufficient maintenance window to perform an upgrade. Depending on which components you are upgrading, the number of hosts in your cluster, and the type of hardware, you might need up to a full day to upgrade your cluster. Before you begin the upgrade, you need to gather some information; these steps are also detailed in the Cloudera Manager and CDH upgrade procedures.

Before upgrading, consult the [release notes](#) for Cloudera Manager and CDH to learn about API changes, deprecated features, new features, and incompatible changes. Also check the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) page to make sure that you are using a supported operating system, JDK, database, and other components.

There are three types of upgrades: major, minor, and maintenance:

Major Upgrades

A major upgrade typically has the following characteristics:

- Large changes to functionality and update of Hadoop to a more recent version
- Incompatible changes in data formats
- Significant changes and additions to the user interface in Cloudera Manager

- Database schema changes for Cloudera Manager that are automatically handled by the upgrade process
- Significant down time for the cluster is required.
- Client configurations are redeployed.

You can only upgrade from one major version to the next major version, for example from version 4.8.1 to version 5.8.0. To be ready for the next major version upgrade, Cloudera recommends that you upgrade to a 5.x version.

Minor Upgrades

Minor upgrades upgrade your software to a higher minor version of a major release—for example from version 5.4.x to version 5.8.x—and typically include the following:

- New functionality
- Bug fixes
- Potential database schema changes for Cloudera Manager that are handled automatically

Incompatible changes or changes to data formats are generally not introduced in minor upgrades. [Client configurations](#) are redeployed.

Maintenance Upgrades

Maintenance upgrades are used only to fix critical bugs or address security issues. No new functionality or incompatible changes are introduced.

Overview of Upgrading Cloudera Manager

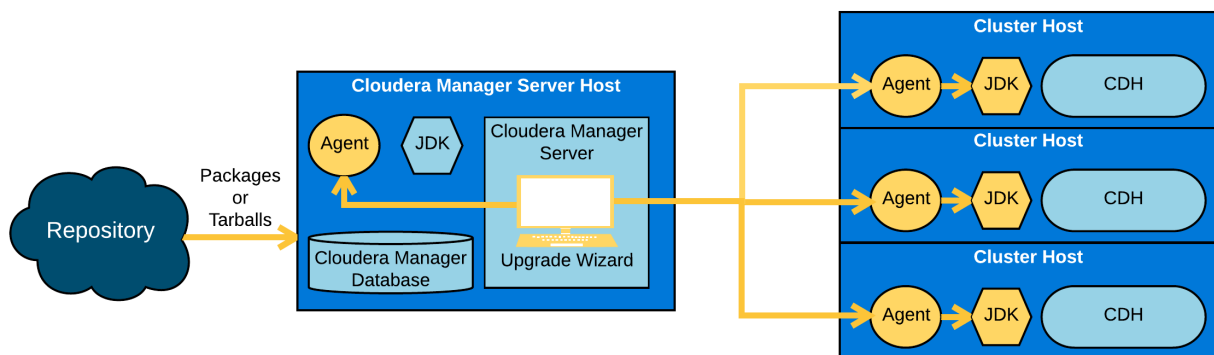


Figure 1: Cloudera Manager 5 Upgrade

To upgrade Cloudera Manager, you perform the following tasks:

1. Upgrade the Cloudera Manager server software on the Cloudera Manager host using operating system package commands from the command line (for example, `yum` on RHEL systems). You can also manually upgrade Cloudera Manager using tarballs. Tarball upgrades are more suitable for environments in which existing tools are used to manage software distributed over many hosts. Cloudera Manager automates much of this process and is recommend for upgrading and managing your CDH clusters.
2. Upgrade the Cloudera Manager agent software on all cluster hosts. The Cloudera Manager upgrade wizard can upgrade the agent software (and, optionally, the JDK), or you can install the agent and JDK software manually from tarballs. The CDH software is not upgraded during this process.

To upgrade Cloudera Manager, see [Upgrading Cloudera Manager](#) on page 12.

Overview of Upgrading CDH

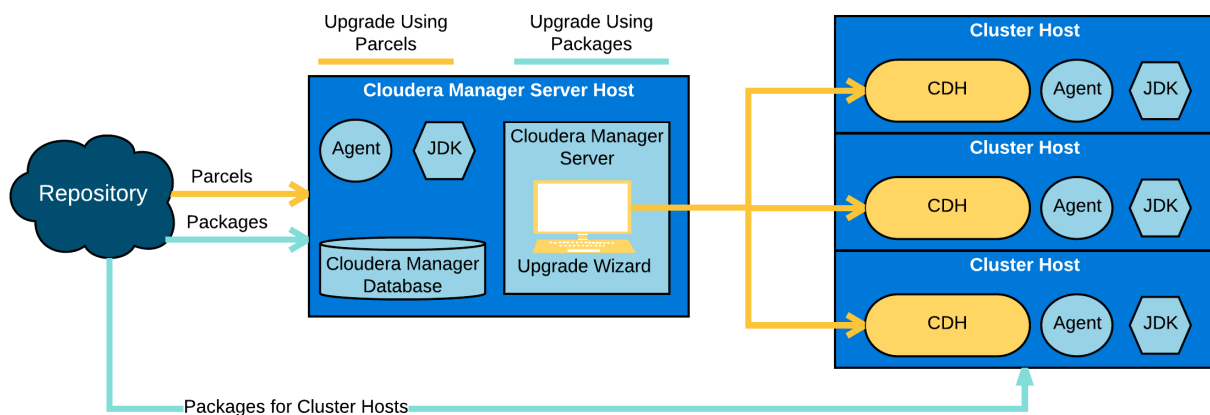


Figure 2: CDH Upgrades

CDH upgrades contain updated versions of the Hadoop software and other components. You can use Cloudera Manager to upgrade CDH using either [parcels or packages](#).

- Upgrade Using Parcels

(Also applies to rolling upgrades.)

Upgrading CDH using parcels is the preferred method because parcels are managed by Cloudera Manager, which automatically downloads, distributes, and activates the correct versions of the software. There are two types of parcel upgrades:

- **Parcels** – Requires you to restart the cluster to complete the upgrade.
- **Rolling Upgrade**– If you have enabled high availability for HDFS, you can perform a rolling upgrade to upgrade CDH without cluster down time. For an easier upgrade experience, consider [switching from packages to parcels](#) so that Cloudera Manager can automate more of the process. You can also switch from packages to parcels when upgrading CDH 5.

- Upgrade Using Packages

Upgrading CDH using packages requires you to download updated packages and manually run package upgrade commands on the Cloudera Manager server and *all cluster hosts*.

See [Upgrading to CDH 5.x Using Packages](#) on page 68.

Cloudera Manager 5.3 introduced an enhanced CDH upgrade wizard that supports major (CDH 4 to CDH 5), minor (CDH 5.x to 5.y), and maintenance upgrades (CDH *a.b.x* to CDH *a.b.y*). Both [parcels and package](#) installations are supported, but packages must be manually installed, whereas parcels are installed by Cloudera Manager.

See [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54

Overview of Upgrading Cloudera Navigator Components

Cloudera Navigator Metadata and Audit servers are automatically upgraded when you upgrade Cloudera Manager. You can also optionally upgrade other Cloudera Navigator components such as Cloudera Navigator Key Trustee Server, Cloudera Navigator Key HSM, and Cloudera Navigator Encrypt. You do not have to upgrade these components along with Cloudera Manager or CDH upgrades.

See [Upgrading Cloudera Navigator Components](#) on page 135.

Upgrading the JDK


Before upgrading Cloudera Manager or CDH, ensure that all cluster hosts are using a [supported version of the Oracle Java Development Kit \(JDK\)](#). All cluster hosts must use the same version of the JDK. See:

- [Upgrading to Oracle JDK 1.7](#) on page 130
- [Upgrading to Oracle JDK 1.8](#) on page 132

Upgrading Cloudera Manager

The procedures for upgrading Cloudera Manager differ depending on the version of Cloudera Manager from which you are upgrading and whether you installed Cloudera Manager using packages or tarballs. See the following links for detailed upgrade instructions:

- Upgrading Cloudera Manager 5
 - [Upgrading Cloudera Manager 5 Using Packages](#) on page 12
Upgrade Cloudera Manager from any version of Cloudera Manager 5 to a higher version of Cloudera Manager 5 using operating system package-management commands.
 - [Upgrading Cloudera Manager 5 Using Tarballs](#) on page 25
Upgrade Cloudera Manager from any version of Cloudera Manager 5 to a higher version of Cloudera Manager 5 manually using tarballs. Recommended for advanced users only.
- [Upgrading Cloudera Manager 4 to Cloudera Manager 5](#) on page 32
Upgrade Cloudera Manager from Cloudera Manager 4 to any version of Cloudera Manager 5 using operating system package-management commands or tarballs. Recommended for advanced users only.

**Note:** Upgrading Cloudera Manager does not upgrade CDH. See [Cloudera Upgrade Overview](#) on page 8 for more complete information about the upgrade process.

Upgrading Cloudera Manager 5 Using Packages

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

This topic describes how to upgrade Cloudera Manager 5.x using *packages*. These steps apply to both [minor and maintenance](#) upgrades. When you upgrade Cloudera Manager using packages, you run operating system package commands from the command line, and then complete the upgrade using Cloudera Manager.

You can also manually upgrade Cloudera Manager using [tarballs](#).

In most cases, you can upgrade Cloudera Manager without shutting down most CDH services. Depending on which services are deployed in your cluster, you might need to stop some dependent services. CDH daemons can run unaffected while Cloudera Manager is upgraded.

The upgrade process does not affect your CDH installation. After upgrading Cloudera Manager, you might want to upgrade CDH 4 clusters to CDH 5, or upgrade to a more recent minor version of CDH.

The Cloudera Manager upgrade process does the following:

- Upgrades the database schema to reflect the current version.
- Upgrades the Cloudera Manager Server and all supporting services.
- Upgrades the Cloudera Manager agent.
- Redeploys client configurations to ensure that client services have the most current configuration.

Steps to upgrade Cloudera Manager 5 using packages:

Step	Description	Link
1	Collect the information you need to upgrade Cloudera Manager. This includes user accounts, passwords, database URLs, and other items. You must gather this information before beginning the upgrade because	Step 1: Collect Upgrade Information on page 13

Step	Description	Link
	some information is available only from the Cloudera Manager Admin Console, which is not available during the upgrade process.	
2	Complete the pre-upgrade steps and review special warnings about upgrades.	Step 2: Complete Pre-Upgrade Steps on page 14
3	Back up the Cloudera Manager databases.	Step 3: Back Up Cloudera Manager Databases on page 14
4	If your Cloudera Manager hosts use an unsupported version of the JDK, you must upgrade the hosts to a supported version of the JDK before upgrading Cloudera Manager. If you plan to upgrade CDH, you must also upgrade the JDK on all cluster hosts.	Step 4: Upgrading to Oracle JDK 1.7 on page 130 Upgrading to Oracle JDK 1.8 on page 132
5	If the Cloudera Manager host does not have access to the internet, or you install a version lower than the latest version of Cloudera Manager, configure access to the Cloudera Manager software from either the Cloudera public repository or a local package repository that you create.	Step 5: Creating and Using a Package Repository for Cloudera Manager on page 157
6	If you are upgrading from Cloudera Navigator 2.6 or lower, upgrade the Cloudera Navigator data management component.	Step 6: Prepare the Cloudera Navigator Data Management Component for Upgrade on page 15
7	Upgrade the Cloudera Manager server and agent software.	Step 7: Upgrade the Cloudera Manager Server on page 16
8	Verify and test your upgrade.	Step 8: Verify and Test the Upgrade on page 24
9	Upgrade any required Cloudera Navigator components: <ul style="list-style-type: none"> • Cloudera Manager Key Trustee Server • Cloudera Navigator Key HSM • Cloudera Navigator Key Trustee KMS • Cloudera Navigator Encrypt. <p>The Cloudera Navigator Data Management Component is upgraded automatically when you upgrade Cloudera Manager.</p>	Step 9: Upgrading Cloudera Navigator Components on page 135
10	(Optional) Upgrade CDH. You are not required to upgrade CDH after upgrading Cloudera Manager. You can upgrade CDH at a later time.	Step 10: (Optional) Upgrade CDH on page 25

If you encounter problems during an upgrade, see the following:

- [Troubleshooting Installation and Upgrade Problems](#)
- [Re-Running the Cloudera Manager Upgrade Wizard](#) on page 163
- [Reverting a Failed Cloudera Manager Upgrade](#) on page 164

Step 1: Collect Upgrade Information

Before starting an upgrade, collect the following information:

1. Host credentials. You must have SSH access and be able to log in using a root account or an account that has password-less sudo permission.
2. The version of Cloudera Manager used in your cluster. Go to **Support > About**.
3. The version of the JDK deployed in the cluster. Go to **Support > About**.
4. The version of CDH. The CDH version number displays next to the cluster name on the **Home** page.

5. Whether the cluster was installed using parcels or packages. This information displays next to the CDH version on the **Home** page of Cloudera Manager.
6. The services enabled in your cluster. Go to **Clusters > Cluster name**.
7. Operating system type and version. Go to **Hosts** and click on a hostname in the list. The operating system type and version displays in the **Distribution** row in the **Details** section.

Step 2: Complete Pre-Upgrade Steps

Before beginning a Cloudera Manager upgrade, do the following:

1. Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) for the new versions you are upgrading to.
2. Read the [Cloudera Manager 5 Release Notes](#).
3. Read the [Cloudera Security Bulletins](#).
4. Note the following:
 - **Cloudera Management Service TLS/SSL configuration**
If you have enabled TLS security for the Cloudera Manager Admin Console, as of Cloudera Manager 5.1, Cloudera Management Service roles communicate with Cloudera Manager using TLS, and fail to start until TLS/SSL properties have been configured.
 - **Navigator**
If you have enabled auditing with Cloudera Navigator, during the upgrade to Cloudera Manager 5, auditing is suspended and is only restarted when you restart the roles of audited services. You will be instructed to stop some services in a later step.
5. If you have previously installed Kafka 1.2, and are upgrading from Cloudera Manager 5.4 or lower, remove the Kafka CSD:
 - a. Determine the location of the CSD directory:
 - a. Select **Administration > Settings**.
 - b. Click the **Custom Service Descriptors** category.
 - c. Retrieve the directory from the **Local Descriptor Repository Path** property.
 - b. Delete the Kafka CSD from the directory.
6. Review package (RPM) dependencies. A Cloudera Manager upgrade might introduce new package dependencies. If your organization has restrictions or requires prior approval for installation of packages, see the list of [Package Dependencies](#) on page 52 before upgrading Cloudera Manager.

Step 3: Back Up Cloudera Manager Databases

1. Stop the Cloudera Management Service:
 - a. Select **Clusters > Cloudera Management Service**.
 - b. Select **Actions > Stop**.
2. Back up the following Cloudera Manager databases:
 - Cloudera Manager Server
 - Cloudera Navigator Audit Server
 - Cloudera Navigator Metadata Server
 - Activity Monitor
 - Reports Manager

To locate information about these databases (database type, hostname, and credentials):

- Cloudera Manager Server – Log in to the Cloudera Manager host and examine the `/etc/cloudera-scm-server/db.properties` file. For example:

```
more /etc/cloudera-scm-server/db.properties
# Auto-generated by scm_prepare_database.sh on Fri Dec 9 08:51:29 PST 2016
#
# For information describing how to configure the Cloudera Manager Server
# to connect to databases, see the "Cloudera Manager Installation Guide."
#
com.cloudera.cmf.db.type=mysql
com.cloudera.cmf.db.host=localhost
com.cloudera.cmf.db.name=cm
com.cloudera.cmf.db.user=cm
com.cloudera.cmf.db.password=cm
```

- For the other databases, go to **Clusters > Cloudera Management Service > Configuration** and select the **Database** category. You might need to contact your database administrator to obtain passwords.

See [Backing Up Databases](#) for detailed instructions for each supported type of database.

3. Start the Cloudera Management Service:

- a. Select **Clusters > Cloudera Management Service**.
- b. Select **Actions > Start**.

Step 4: Upgrade the JDK

If your Cloudera Manager hosts use an unsupported version of the JDK, you must upgrade the hosts to a [supported version of the JDK](#) before upgrading Cloudera Manager. If you plan to upgrade CDH, you must also upgrade the JDK on all cluster hosts.

See:

- [Upgrading to Oracle JDK 1.7](#) on page 130
- [Upgrading to Oracle JDK 1.8](#) on page 132

If you have enabled TLS/SSL, you must reinstall CA certificates to your truststores after upgrading the JDK. See [Creating and Using the Truststore](#).

Step 5: Establish Access to the Software

If the Cloudera Manager host does not have access to the internet, or you install a version lower than the latest version of Cloudera Manager, configure access to the Cloudera Manager software from either the Cloudera public repository or a local package repository that you create.

See [Creating and Using a Package Repository for Cloudera Manager](#) on page 157.


Step 6: Prepare the Cloudera Navigator Data Management Component for Upgrade

Cloudera Manager upgrades Cloudera Navigator as part of the Cloudera Manager upgrade process. If you are upgrading from Cloudera Navigator 2.6 or lower, follow the steps in this section to prepare the Cloudera Navigator data management component for upgrade and then continue with [Step 7: Upgrade the Cloudera Manager Server](#) on page 16. If you are upgrading from Cloudera Navigator 2.7 or higher, skip this section and continue with [Step 7: Upgrade the Cloudera Manager Server](#) on page 16.

To find the version of Cloudera Navigator:

1. Go to **Clusters > Cloudera Navigator**.

The Cloudera Navigator user interface displays.

2. Log in to Cloudera Navigator.
3. Click the  icon and select **About**.

A dialog box displays the version number and other information about Cloudera Navigator.

To upgrade the Cloudera Navigator data management component:

1. Stop the Navigator Metadata Server role:
 - a. Go to **Clusters > Cloudera Management Service > Instances**.
 - b. Select the **Navigator Metadata Server**.
 - c. Click **Actions for Selected > Stop**.
2. Back up the Navigator Metadata Server storage directory. To find the location of this directory:
 - a. Go to **Clusters > Cloudera Management Service > Instances**.
 - b. Click the **Configuration** tab.
 - c. Select **Scope > Navigator Metadata Server**.

The **Navigator Metadata Server Storage Dir** property stores the location of the directory.
3. Start the Navigator Metadata Server role.
 - a. Go to **Clusters > Cloudera Management Service > Instances**.
 - b. Select the **Navigator Metadata Server**.
 - c. Click **Actions for Selected > Start**.
4. Purge the Navigator Metadata. See [Purging Metadata for HDFS Entities, Hive and Impala Select Queries, and YARN, Sqoop, and Pig Operations](#).
5. Stop the Navigator Metadata Server role:
 - a. Go to **Clusters > Cloudera Management Service > Instances**.
 - b. Select the **Navigator Metadata Server**.
 - c. Click **Actions for Selected > Stop**.
6. [Make sure that the Navigator Metadata Server has sufficient memory](#) to complete the upgrade.
7. If you are using an Oracle database, in SQL*Plus, ensure that the following additional privileges are set:

```
GRANT EXECUTE ON sys.dbms_crypto TO nav;  
GRANT CREATE VIEW TO nav;
```

where *nav* is the user of the Navigator Audit Server database.

Step 7: Upgrade the Cloudera Manager Server

1. If your cluster is running the embedded PostgreSQL database, stop all services that are using the embedded database. These can include:
 - Hive service and all services such as Impala and Hue that use the Hive metastore
 - Oozie
 - Sentry
2. If your cluster is running the Cloudera Navigator data management component and the following services are enabled for auditing, [stop](#) the following roles. (You also can elect to skip this step and leave these services running, but some audits by Cloudera Navigator may not occur during the Cloudera Manager upgrade process.)
 - **HDFS** - NameNode
 - **HBase** - Master and RegionServers
 - **Hive** - HiveServer2
 - **Hue** - Beeswax Server



Note: Stopping these roles renders any service depending on these roles unavailable. For the HDFS - NameNode role, most of the services in the cluster are unavailable until the upgrade is finished.

To determine which services are enabled for auditing:

- a. Go to the **Home** page in Cloudera Manager.
- b. Click **Configuration > Navigator Settings**.
- c. Type "Enable Audit" in the search box. The **Enable Audit Collection** property displays the services for which Cloudera Navigator auditing is enabled.

3. Stop the Cloudera Management Service:

- a. Select **Clusters > Cloudera Management Service**.
- b. Select **Actions > Stop**.

4. Stop Cloudera Manager Server, Database, and Agent:

- a. Use the Cloudera Manager Admin Console to stop any running commands. These include user commands and commands Cloudera Manager automatically triggers in response to a state change or a schedule. You can either wait for commands to complete, or stop any running commands. For more information on viewing and stopping running commands, see [Viewing Running and Recent Commands](#).



Important: If you do not stop all commands, the Cloudera Manager Server *fails to start after upgrade*.

- b. On the host running the Cloudera Manager Server, stop the Cloudera Manager Server:

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

- c. If you are using the embedded PostgreSQL database with Cloudera Manager, stop the database on the host where the database runs, usually the Cloudera Manager Server host:

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



Important:

If you are *not* running the embedded database service and you attempt to stop it, you receive a message indicating that the service cannot be found. If instead you get a message that the shutdown failed, the embedded database is still running, probably because services are connected to the Hive metastore. If the database shutdown fails due to connected services, issue the following command:

- RHEL-compatible 7 and higher:

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

- All other Linux distributions:

```
sudo service cloudera-scm-server-db fast_stop
```

- d. If the Cloudera Manager host is also running the Cloudera Manager Agent, stop the Cloudera Manager Agent:

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

5. Back up the following directories on the Cloudera Manager server host:

- /etc/cloudera-scm-server
- /etc/cloudera-scm-agent

6. Establish access to the Cloudera Manager Server packages. You can either upgrade from the Cloudera repository at <https://archive.cloudera.com/cm5/>, or you can create your own package repository, as described in [Creating and Using a Package Repository for Cloudera Manager](#) on page 157. You must create your own repository if Cloudera Manager does not have Internet access or you want to upgrade to a version of Cloudera Manager lower than the latest version.

To upgrade using the Cloudera repository:

- a. Back up the current Cloudera Manager repo file, located in one of the following directories:

RHEL

`/etc/yum.repos.d/`

SLES

`/etc/zypp/repos.d/`

Ubuntu or Debian

`/etc/apt/sources.list.d/`

- b. Download the Cloudera `.repo` file for your distribution by starting at <https://archive.cloudera.com/cm5/> and navigating to the directory that matches your operating system.
 - For Red Hat or CentOS 6, go to the appropriate release directory, for example, https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/. In that directory, find the repo file that contains information including the repository base URL and GPG key. The contents of the `cloudera-manager.repo` are similar to the following:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 5, on RHEL or CentOS 6 x86_64
name=Cloudera Manager
baseurl=https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5/
gpgkey = https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

- For Ubuntu or Debian systems, go to the appropriate release directory, for example, <https://archive.cloudera.com/cm4/debian/wheezy/amd64/cm>. The repo file, in this case, `cloudera.list`, is similar to the following:

```
# Packages for Cloudera Manager, Version 5, on Debian 7.0 x86_64
deb https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
deb-src https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
```

For example, you can use the following command to download the `.repo` file for Cloudera Manager version 5 and RHEL version 6:

```
$ wget https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/cloudera-manager.repo
```

- c. Do one of the following on the Cloudera Manager Server host:
 - **Upgrade to the most recent version of Cloudera Manager:**

Copy the `cloudera-manager.repo` file to the configuration location for the package management software for your system:

RHEL

Copy `cloudera-manager.repo` to `/etc/yum.repos.d/`

SLES

Copy `cloudera-manager.repo` to `/etc/zypp/repos.d/`

Ubuntu or Debian

Copy `cloudera.list` to `/etc/apt/sources.list.d/`

- **Upgrade to an specific version of Cloudera Manager:**

RHEL-compatible or SLES

1. Edit the `cloudera-manager.repo` file to change the `baseurl` to point to the version of Cloudera Manager you want to download. For example, to install Cloudera Manager version 5.0.1, change:

```
baseurl=https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5/
```

to:

```
baseurl=https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5.0.1/.
```

2. Save the edited file:

- For RHEL or CentOS, save it in `/etc/yum.repos.d/`.
- For SLES, save it in `/etc/zypp/repos.d/`.

Ubuntu or Debian

1. Download the Cloudera Manager list file (`cloudera.list`) using the links provided at [Cloudera Manager Version and Download Information](#). For example, for Ubuntu 10.04 (lucid), this file is located at

```
https://archive.cloudera.com/cm5/ubuntu/lucid/amd64/cm/cloudera.list.
```

2. Edit the `cloudera.list` file to change the second-to-last element to specify the version of Cloudera Manager you want to install. For example, with Ubuntu lucid, if you want to install Cloudera Manager version 5.0.1, change:

```
deb https://archive.cloudera.com/cm5/ubuntu/lucid/amd64/cm lucid-cm5 contrib
```

to:

```
deb https://archive.cloudera.com/cm5/ubuntu/lucid/amd64/cm lucid-cm5.0.1
contrib.
```

3. Save the edited file in the `/etc/apt/sources.list.d/` directory.

- d. Run the following command to clean the cache directories and upgrade the software:

RHEL

```
sudo yum clean all
sudo yum upgrade cloudera-manager-server cloudera-manager-daemons cloudera-manager-agent
```



Note:

- `yum clean all` cleans yum cache directories, ensuring that you download and install the latest versions of the packages.
- If your system is not up to date, any underlying system components must be upgraded before `yum update` can run. `yum` indicates which components must be upgraded.
- If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add `cloudera-manager-server-db-2` to the list of packages in the `yum upgrade` command. *The embedded PostgreSQL database should not be used in production environments.*

SLES

```
sudo zypper clean --all
sudo zypper up -r https://archive.cloudera.com/cm5/sles/11/x86_64/cm/5/
```

To upgrade 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
```

Ubuntu or Debian

The following commands clean cached repository information and update Cloudera Manager components:

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

If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add `cloudera-manager-server-db-2` to the list of packages in the `apt-get install` command. *The embedded PostgreSQL database should not be used in production environments.*

During this process, you might be prompted about 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 prompts.

7. If you customized the `/etc/cloudera-scm-agent/config.ini` file, your customized file is renamed with the extension `.rpmsave` or `.dpkg-old`. Merge any customizations into the `/etc/cloudera-scm-agent/config.ini` file that is installed by the package manager.
8. On the Cloudera Manager Server host, verify that you now have the following packages, corresponding to the version of Cloudera Manager you installed, by running the following command:

RPM-based distributions

```
$ rpm -qa 'cloudera-manager-*'
cloudera-manager-server-5.11.0-0.cm5110.p0.97.el6.x86_64
cloudera-manager-agent-5.11.0-0.cm5110.p0.97.el6.x86_64
cloudera-manager-daemons-5.11.0-0.cm5110.p0.97.el6.x86_64
```

Ubuntu or Debian

```
~# 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-agent 5.11.0-0.cm5110.p0.97~sq The Cloudera Manager Agent
ii cloudera-manager-daemo 5.11.0-0.cm5110.p0.97~sq Provides daemons for monitoring
Hadoop and related tools.
ii cloudera-manager-serve 5.11.0-0.cm5110.p0.97~sq The Cloudera Manager Server
```



Note: You might also see an entry for the `cloudera-manager-server-db-2` if you are using the embedded PostgreSQL database, and additional packages for plug-ins, depending on what was previously installed on the server host. If the `cloudera-manager-server-db-2` package is installed, and you do not plan to use the embedded database, you can remove this package.

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

- a. If you are using the embedded PostgreSQL database for Cloudera Manager, start the database. If your installation uses other databases, Cloudera Manager reconnects with them after start up.

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

- b. Start the Cloudera Manager Server:

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

You should see the following:

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

10 Log in to the Cloudera Manager Admin Console. It can take several minutes for Cloudera Manager Server to start, and the console is unavailable until the server startup is complete.

The **Upgrade Wizard** displays.

11 Upgrade the Cloudera Manager Agent [using Cloudera Manager](#) or by [manually upgrading the packages](#):

Cloudera Manager upgrades Agent software

When Cloudera Manager upgrades the Cloudera Manager agent, Cloudera Manager handles the upgrade and cleanup, and optionally upgrades the JDK.

1. Select **Yes, I would like to upgrade the Cloudera Manager Agent packages now** and click **Continue**.
2. Select the release of the Cloudera Manager Agent to install. Normally, this is the **Matched Release for this Cloudera Manager Server**. However, if you used a custom repository (instead of `archive.cloudera.com`) for the Cloudera Manager server, select **Custom Repository** and provide the required information. The custom repository location must contain the matched Agent version. See [Creating and Using a Package Repository for Cloudera Manager](#) on page 157. Custom repositories are required when your cluster cannot access the Internet, or when you need to upgrade to a version other than the most current version of Cloudera Manager.
3. Click **Continue**. The JDK Installation Options page displays.
 - If you want Cloudera Manager to install JDK 1.7 on all cluster hosts, select **Install Oracle Java SE Development Kit (JDK)**.
 - If local laws permit you to deploy unlimited strength encryption, and you are running a secure cluster, select the **Install Java Unlimited Strength Encryption Policy Files** checkbox.
4. Click **Continue**.
5. Specify credentials and initiate Agent installation:
 - Select **root** or enter the username for an account that has password-less sudo permission.
 - Select an authentication method:
 - If you choose password authentication, enter and confirm the password.
 - If you choose public-key authentication, provide a passphrase and path to the required key files.
 - You can 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.

6. Click **Continue**.

The Cloudera Manager Agent packages and, if selected, the JDK are installed.

7. Click **Continue**.

The Host Inspector runs to inspect your managed hosts for correct versions and configurations. If problems occur, you can make changes and then rerun the inspector.

When you are satisfied with the inspection results, click **Continue**.

Manually upgrade Agent software

To manually upgrade the Cloudera Manager Agent software, you use package commands to clean up old versions, download the new version, and upgrade the software.

To manually upgrade the Cloudera Manager agent:

1. On all cluster hosts except the Cloudera Manager Server host, stop the Agent:

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

2. Select **No, I would like to skip the agent upgrade now** and click **Continue**.
3. Copy the repo file as described in step 6 on page 18.
4. Run the following commands on all hosts except the Cloudera Manager Server host:

RHEL

```
sudo yum clean all
sudo yum upgrade cloudera-manager-daemons cloudera-manager-agent
```



Note:

- `yum clean all` cleans `yum` cache directories, ensuring that you download and install the latest versions of the packages.
- If your system is not up to date, any underlying system components must be upgraded before `yum update` can run. `yum` indicates which components must be upgraded.
- If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add `cloudera-manager-server-db-2` to the list of packages in the `yum` upgrade command. *Do not use the embedded PostgreSQL database in production environments.*

SLES

```
sudo zypper clean --all
sudo zypper up -r https://archive.cloudera.com/cm5/sles/11/x86_64/cm/5/
```

To upgrade 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
```

Ubuntu or Debian

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 dist-upgrade
sudo apt-get install cloudera-manager-agent cloudera-manager-daemons
```

If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add `cloudera-manager-server-db-2` to the list of packages in the `apt-get install` command. *Do not use the embedded PostgreSQL database in production environments.*

During this process, you might be prompted about 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 prompts.

5. If you customized the `/etc/cloudera-scm-agent/config.ini` file, your customized file is renamed with the extension `.rpmsave` or `.dpkg-old`. Merge any customizations into the `/etc/cloudera-scm-agent/config.ini` file that is installed by the package manager.
6. On all cluster hosts, start the Agent:

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

7. Click **Continue**. The Host Inspector inspects your managed hosts for correct versions and configurations. If problems occur, you can make changes and then rerun the inspector. When you are satisfied with the inspection results, click **Continue**.

12 Click Finish.

- 13 If you are upgrading from Cloudera Manager 5.0 and are using an external database for Cloudera Navigator, the Database Setup page displays. Configure these database settings:

- a. Enter the database host, database type, database name, username, and password for the database.
- b. Click **Test Connection** to confirm that Cloudera Manager can communicate with the Cloudera Navigator database using the information you supplied. If the test succeeds in all cases, click **Continue**; otherwise, check and correct the information you have provided for the database and then try the test again. (For some servers, if you are using the embedded database, you will see a message saying the database will be created in a later step.)

- 14 The **Review Changes** page displays. Review the configuration changes to be applied and click **Continue**. The Upgrade wizard displays a dialog box allowing you to choose whether to restart the Cloudera Management Service.

15 Click Continue.

If you keep the default selection, the Upgrade wizard restarts the Cloudera Management Service.

16 Click Finish.

The **Home** page displays.

All services (except for any services you stopped) should now be running.

- 17 If, as part of this upgrade, you stopped some selected services and roles, [restart](#) the following roles:

- **HDFS** - NameNode
- **HBase** - Master and RegionServers
- **Hive** - HiveServer2
- **Hue** - Beeswax Server

18 If you upgraded the JDK, do the following:

- a. If the Cloudera Manager Server host is also running a Cloudera Manager Agent, restart the Cloudera Manager Server:

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

If the Cloudera Manager Server does not start, see [Troubleshooting Installation and Upgrade Problems](#).

- b. Restart all services:

- a. On the **Home > Status** tab, click



next to the cluster name and select **Restart**.

- b. In the confirmation dialog box that displays, click **Restart**.

19 If Cloudera Manager reports [stale configurations](#) after the upgrade, restart the cluster services and redeploy the client configurations. If you are also upgrading CDH, this step is not required.

- a. On the **Home > Status** tab, click



next to the cluster name and select **Restart**.

- b. In the confirmation dialog box, click **Restart**.

- c. On the **Home > Status** tab, click



next to the cluster name and select **Deploy Client Configuration**.

- d. In the confirmation dialog box, click **Deploy Client Configuration**.

20 If upgrading from Navigator 2.6 or lower:

- a. [Start and log into the Cloudera Navigator data management component UI](#). The Upgrading Navigator page displays. Depending on the amount of data in the Navigator Metadata Server storage directory, the upgrade process can take *three to four hours* or longer.

21 When the upgrade is complete, click **Continue**. The Cloudera Navigator landing page is displayed.

22 If you are upgrading from CDH 5.5.0 or lower to CDH 5.5.0 or higher, hard restart the agent on all hosts to update and restart the `supervisord` process:

RHEL 7 and higher:

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

Other Linux Distributions:

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

Step 8: Verify and Test the Upgrade

If the commands to update and start the Cloudera Manager Server complete without errors, the upgrade has completed successfully. To verify, check that the server versions have been updated.

1. Verify that the agents are sending heartbeats to Cloudera Manager:

- a. Go to **Hosts > All Hosts**.
- b. Click the column header labeled **Last Heartbeat** to sort it.
- c. Verify that the last heartbeat for each host has occurred within one minute.

2. In the Cloudera Manager Admin Console, click the **Hosts** tab.
3. Click **Inspect All Hosts**. On large clusters, the host inspector can take some time to finish running. You must wait for the process to complete before proceeding to the next step.
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.
5. Verify that the monitoring features are working as expected; follow the instructions in [Testing the Installation](#).

Step 9: Upgrade any Required Navigator Components

Upgrade any Cloudera Navigator components deployed in your cluster:

- Cloudera Manager Key Trustee Server
- Cloudera Navigator Key HSM
- Cloudera Navigator Key Trustee KMS
- Cloudera Navigator Encrypt.

You can upgrade other Cloudera Navigator components at any time. You do not have to perform these upgrades when upgrading Cloudera Manager or CDH.

See [Upgrading Cloudera Navigator Components](#) on page 135.

Step 10: (Optional) Upgrade CDH

Cloudera Manager 5 can manage both CDH 4 and CDH 5, so upgrading existing CDH 4 and CDH 5 installations is not required. To upgrade CDH, see [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54.

Upgrading Cloudera Manager 5 Using Tarballs

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

This topic describes how to upgrade Cloudera Manager 5.x using *tarballs*. Tarballs contain both the Cloudera Manager Server and Cloudera Manager Agent in a single file.

In most cases it is possible to upgrade Cloudera Manager without shutting down most CDH services, although you may need to stop some dependent services. CDH daemons can continue running, unaffected, while Cloudera Manager is upgraded. The upgrade process does not affect your CDH installation. After upgrading Cloudera Manager you may also want to upgrade CDH 4 clusters to CDH 5, or upgrade to a more recent minor version of CDH.

Steps to upgrade Cloudera Manager 5 using tarballs:

Step	Description	Link
1	Collect the information you need to upgrade Cloudera Manager. This includes user accounts, passwords, database URLs, and other items. You must gather this information before beginning the upgrade because some information is available only from the Cloudera Manager Admin Console, which is not available during the upgrade process.	Step 1: Collect Upgrade Information on page 26
2	Complete the pre-upgrade steps and review special warnings about upgrades.	Step 2: Complete Pre-Upgrade Steps on page 26
3	If your Cloudera Manager hosts use an unsupported version of the JDK, you must upgrade the hosts to a supported version of the JDK before upgrading Cloudera Manager. If you plan to upgrade CDH, you must also upgrade the JDK on all cluster hosts.	Step 3: Upgrade the JDK Upgrading to Oracle JDK 1.7 on page 130 Upgrading to Oracle JDK 1.8 on page 132

Step	Description	Link
4	If the Cloudera Manager host does not have access to the internet, or you install a version lower than the latest version of Cloudera Manager, configure access to the Cloudera Manager software from either the Cloudera public repository or a local package repository that you create.	Step 4: Creating and Using a Package Repository for Cloudera Manager on page 157
5	If you are upgrading from Cloudera Navigator 2.6 or lower, upgrade the Cloudera Navigator data management component.	Step 5: Upgrading the Cloudera Navigator Data Management Component on page 135
6	Upgrade the Cloudera Manager server and agent software.	Step 6: Upgrading the Cloudera Manager Server and Agents on page 27
7	Verify and test your upgrade.	Step 7: Verify and Test the Upgrade on page 31
8	Upgrade any required Cloudera Navigator components: <ul style="list-style-type: none"> • Cloudera Manager Key Trustee Server • Cloudera Navigator Key HSM • Cloudera Navigator Key Trustee KMS • Cloudera Navigator Encrypt. The Cloudera Navigator Data Management Component is upgraded automatically when you upgrade Cloudera Manager.	Step 8: Upgrading Cloudera Navigator Components on page 135
9	(Optional) Upgrade CDH. You are not required to upgrade CDH after upgrading Cloudera Manager. You can upgrade CDH at a later time.	Step 9: (Optional) Upgrade CDH on page 32

Step 1: Collect Upgrade Information

Before starting an upgrade, collect the following information:

1. Host credentials. You must have SSH access and be able to log in using a root account or an account that has password-less sudo permission.
2. The version of Cloudera Manager used in your cluster. Go to **Support > About**.
3. The version of the JDK deployed in the cluster. Go to **Support > About**.
4. The version of CDH. The CDH version number displays next to the cluster name on the **Home** page.
5. Whether the cluster was installed using parcels or packages. This information displays next to the CDH version on the **Home** page of Cloudera Manager.
6. The services enabled in your cluster. Go to **Clusters > Cluster name**.
7. Operating system type and version. Go to **Hosts** and click on a hostname in the list. The operating system type and version displays in the **Distribution** row in the **Details** section.

Step 2: Complete Pre-Upgrade Steps

Before beginning a Cloudera Manager upgrade, do the following:

1. Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) for the new versions you are upgrading to.
2. Read the [Cloudera Manager 5 Release Notes](#).
3. Read the [Cloudera Security Bulletins](#).
4. Note the following:
 - **Cloudera Management Service TLS/SSL configuration**

If you have enabled TLS security for the Cloudera Manager Admin Console, as of Cloudera Manager 5.1, Cloudera Management Service roles communicate with Cloudera Manager using TLS, and fail to start until TLS/SSL properties have been configured.

- **Navigator**

If you have enabled auditing with Cloudera Navigator, during the upgrade to Cloudera Manager 5, auditing is suspended and is only restarted when you restart the roles of audited services. You will be instructed to stop some services in a later step.

5. If you have previously installed Kafka 1.2, and are upgrading from Cloudera Manager 5.4 or lower, remove the Kafka CSD:
 - a. Determine the location of the CSD directory:
 - a. Select **Administration > Settings**.
 - b. Click the **Custom Service Descriptors** category.
 - c. Retrieve the directory from the **Local Descriptor Repository Path** property.
 - b. Delete the Kafka CSD from the directory.
6. Review package (RPM) dependencies. A Cloudera Manager upgrade might introduce new package dependencies. If your organization has restrictions or requires prior approval for installation of packages, see the list of [Package Dependencies](#) on page 52 before upgrading Cloudera Manager.

Step 3: Upgrade the JDK

See:

- [Upgrading to Oracle JDK 1.7](#) on page 130
- [Upgrading to Oracle JDK 1.8](#) on page 132

Step 4: Creating and Using a Package Repository for Cloudera Manager

If the Cloudera Manager host does not have access to the internet, or you install a version lower than the latest version of Cloudera Manager, configure access to the Cloudera Manager software from either the Cloudera public repository or a local package repository that you create.

See: [Creating and Using a Package Repository for Cloudera Manager](#) on page 157

Step 5: Upgrading the Cloudera Navigator Data Management Component

If you are upgrading from Cloudera Navigator 2.6 or lower, upgrade the Cloudera Navigator data management component.

See [Upgrading the Cloudera Navigator Data Management Component](#) on page 135

Step 6: Upgrading the Cloudera Manager Server and Agents

1. If your cluster is running the embedded PostgreSQL database, stop all services that are using the embedded database. These can include:
 - Hive service and all services such as Impala and Hue that use the Hive metastore
 - Oozie
 - Sentry
2. Stop Cloudera Manager Server, Database, and Agent:
 - a. Use the Cloudera Manager Admin Console to stop any running commands. These include user commands and commands Cloudera Manager automatically triggers in response to a state change or a schedule. You can either wait for commands to complete, or stop any running commands. For more information on viewing and stopping running commands, see [Viewing Running and Recent Commands](#).



Important: If you do not stop all commands, the Cloudera Manager Server *fails to start after upgrade*.

b. On the host running the Cloudera Manager Server, stop the Cloudera Manager Server:

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

c. If you are using the embedded PostgreSQL database with Cloudera Manager, stop the database on the host where the database runs, usually the Cloudera Manager Server host:

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



Important:

If you are *not* running the embedded database service and you attempt to stop it, you receive a message indicating that the service cannot be found. If instead you get a message that the shutdown failed, the embedded database is still running, probably because services are connected to the Hive metastore. If the database shutdown fails due to connected services, issue the following command:

- RHEL-compatible 7 and higher:

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

- All other Linux distributions:

```
sudo service cloudera-scm-server-db fast_stop
```

d. If the Cloudera Manager host is also running the Cloudera Manager Agent, stop the Cloudera Manager Agent:

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

3. Download tarballs from the locations listed in [Cloudera Manager Version and Download Information](#).

4. Copy the tarballs and unpack them on all hosts on which you intend to install Cloudera Manager Server and Cloudera Manager Agents, in a directory of your choosing. If necessary, create a new directory to accommodate the files you extract from the tarball. For instance, if `/opt/cloudera-manager` does not exist, create it using a command similar to:

```
$ sudo mkdir /opt/cloudera-manager
```

5. Extract the contents of the tarball to this directory. For example, to copy a tar file to your home directory and extract the contents of all tar files to the `/opt/` directory, use a command similar to the following:

```
$ sudo tar xzf cloudera-manager*.tar.gz -C /opt/cloudera-manager
```

The files are extracted to a subdirectory named according to the Cloudera Manager version being extracted. For example, files could be extracted to `/opt/cloudera-manager/cm-5.0/`. This full path is needed later and is referred to as *tarball_root* directory.

6. On every Cloudera Manager Agent host, configure the Cloudera Manager Agent to point to the Cloudera Manager Server by setting the following properties in the `tarball_root/etc/cloudera-scm-agent/config.ini` configuration file:

Property	Description
<code>server_host</code>	Name of the host where Cloudera Manager Server is running.
<code>server_port</code>	Port on the host where Cloudera Manager Server is running.

7. By default, a tarball installation has a `var` subdirectory where state is stored. In a non-tarball installation, state is stored in `/var`. Cloudera recommends that you configure the tarball installation to use an external directory as the `/var` equivalent (`/var` or any other directory outside the tarball) so that when you upgrade Cloudera Manager, the new tarball installation can access this state. Configure the installation to use an external directory for storing state by editing `tarball_root/etc/default/cloudera-scm-agent` and setting the `CMF_VAR` variable to the location of the `/var` equivalent. If you do not reuse the state directory between different tarball installations, duplicate Cloudera Manager Agent entries can occur in the Cloudera Manager database.

8. Start Cloudera Manager Server.

The way in which you start the Cloudera Manager Server varies depending on which account you want the Server to run under:

- As root:

```
sudo tarball_root/etc/init.d/cloudera-scm-server start
```

- As another user. If you run as another user, ensure the user you created for Cloudera Manager owns the location to which you extracted the tarball including the newly created database files. If you followed the earlier examples and created the directory `/opt/cloudera-manager` and the user `cloudera-scm`, you could use the following command to change ownership of the directory:

```
sudo chown -R cloudera-scm:cloudera-scm /opt/cloudera-manager
```

Once you have established ownership of directory locations, you can start Cloudera Manager Server using the user account you chose. For example, you might run the Cloudera Manager Server as `cloudera-service`. In this case, you have the following options:

- Run the following command:

```
$ sudo -u cloudera-service tarball_root/etc/init.d/cloudera-scm-server start
```

- Edit the configuration files so the script internally changes the user. Then run the script as root:

1. Remove the following line from `tarball_root/etc/default/cloudera-scm-server`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `tarball_root/etc/init.d/cloudera-scm-server` to the user you want the server to run as. For example, to run as `cloudera-service`, change the user and group as follows:

```
USER=cloudera-service
GROUP=cloudera-service
```

3. Run the server script as root:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-server start
```

- To start the Cloudera Manager Server automatically after a reboot:

1. Run the following commands on the Cloudera Manager Server host:

- **RHEL-compatible and SLES**

```
$ cp tarball_root/etc/init.d/cloudera-scm-server /etc/init.d/cloudera-scm-server
$ chkconfig cloudera-scm-server on
```

- **Debian/Ubuntu**

```
$ cp tarball_root/etc/init.d/cloudera-scm-server /etc/init.d/cloudera-scm-server
$ update-rc.d cloudera-scm-server defaults
```

2. On the Cloudera Manager Server host, open the `/etc/init.d/cloudera-scm-server` file and change the value of `CMF_DEFAULTS` from `${CMF_DEFAULTS:-/etc/default}` to `tarball_root/etc/default`.

9. To stop the Cloudera Manager Agent, run this command on each Agent host:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent hard_stop_confirmed
```

If you are running [single user mode](#), stop Cloudera Manager Agent using the user account you chose. For example, if you are running the Cloudera Manager Agent as `cloudera-scm`, you have the following options:

- Run the following command:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent hard_stop_confirmed
```

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from `tarball_root/etc/default/cloudera-scm-agent`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `tarball_root/etc/init.d/cloudera-scm-agent` to the user you want the Agent to run as. For example, to run as `cloudera-scm`, change the user and group as follows:

```
USER=cloudera-scm
GROUP=cloudera-scm
```

3. Run the Agent script as root:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent hard_stop_confirmed
```

- 10 Start the Cloudera Manager Agent according to the account you want the Agent to run under:

- To start the Cloudera Manager Agent, run this command on each Agent host:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent start
```

When the Agent starts, it contacts the Cloudera Manager Server.

- If you are running [single user mode](#), start Cloudera Manager Agent using the user account you chose. For example, to run the Cloudera Manager Agent as `cloudera-scm`, you have the following options:

- Run the following command:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent start
```

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from `tarball_root/etc/default/cloudera-scm-agent`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `tarball_root/etc/init.d/cloudera-scm-agent` to the user you want the Agent to run as. For example, to run as `cloudera-scm`, change the user and group as follows:

```
USER=cloudera-scm
GROUP=cloudera-scm
```

3. Run the Agent script as root:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent start
```

- To start the Cloudera Manager Agents automatically after a reboot:

1. Run the following commands on each Agent host:

- **RHEL-compatible and SLES**

```
$ cp tarball_root/etc/init.d/cloudera-scm-agent /etc/init.d/cloudera-scm-agent
$ chkconfig cloudera-scm-agent on
```

- **Debian/Ubuntu**

```
$ cp tarball_root/etc/init.d/cloudera-scm-agent /etc/init.d/cloudera-scm-agent
$ update-rc.d cloudera-scm-agent defaults
```

2. On each Agent, open the `tarball_root/etc/init.d/cloudera-scm-agent` file and change the value of `CMF_DEFAULTS` from `${CMF_DEFAULTS:-/etc/default}` to `tarball_root/etc/default`.

- 11 Log in to the Cloudera Manager Admin Console.

- 12 Restart all services:

- a. On the **Home > Status** tab, click



next to the cluster name and select **Restart**.

- b. In the confirmation dialog box that displays, click **Restart**.

- 13 Upgrade any required Cloudera Navigator components:

- [Upgrading the Cloudera Navigator Data Management Component](#) on page 135
- [Upgrading Cloudera Navigator Key Trustee Server](#) on page 136
- [Upgrading Cloudera Navigator Key HSM](#) on page 151
- [Upgrading Key Trustee KMS](#) on page 152
- [Upgrading Cloudera Navigator Encrypt](#) on page 153

Step 7: Verify and Test the Upgrade

If the commands to update and start the Cloudera Manager Server complete without errors, the upgrade has completed successfully. To verify, check that the server versions have been updated.

1. Verify that the agents are sending heartbeats to Cloudera Manager:

- a. Go to **Hosts > All Hosts**.
- b. Click the column header labeled **Last Heartbeat** to sort it.

- c. Verify that the last heartbeat for each host has occurred within one minute.
2. In the Cloudera Manager Admin Console, click the **Hosts** tab.
3. Click **Inspect All Hosts**. On large clusters, the host inspector can take some time to finish running. You must wait for the process to complete before proceeding to the next step.
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.
5. Verify that the monitoring features are working as expected; follow the instructions in [Testing the Installation](#).

Step 8: Upgrading Cloudera Navigator Components

Upgrade any required Cloudera Navigator components:

- Cloudera Manager Key Trustee Server
- Cloudera Navigator Key HSM
- Cloudera Navigator Key Trustee KMS
- Cloudera Navigator Encrypt.

See [Upgrading Cloudera Navigator Components](#) on page 135.

The Cloudera Navigator Data Management Component is upgraded automatically when you upgrade Cloudera Manager.

Step 9: (Optional) Upgrade CDH

Cloudera Manager 5 can manage both CDH 4 and CDH 5, so upgrading existing CDH 4 and CDH 5 installations is not required. To upgrade CDH, see [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54.

Upgrading Cloudera Manager 4 to Cloudera Manager 5

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

This process applies to upgrading all versions of Cloudera Manager 4 to Cloudera Manager 5.

In most cases, you can upgrade without shutting down most CDH services, although you may need to stop some dependent services. CDH daemons can run unaffected while Cloudera Manager is upgraded, and the upgrade process does not affect your CDH installation. However, to use Cloudera Manager 5 features, all services must be restarted after the upgrade. After upgrading Cloudera Manager, you may also want to upgrade CDH 4 clusters to CDH 5.

Follow these steps to upgrade Cloudera Manager 4 to the latest version of Cloudera Manager.

Review Warnings and Notes

**Warning:**

- **Cloudera Management Service databases**

Cloudera Manager 5 stores Host and Service Monitor data in a local datastore. The Cloudera Manager 4 to Cloudera Manager 5 upgrade wizard automatically migrates data from existing embedded PostgreSQL or external databases to the local datastore. See [Data Storage for Monitoring Data](#).

The Host Monitor and Service Monitor databases are stored on the partition hosting `/var`. Ensure that you have at least 20 GB available on the partition.

If you have been storing the data in an external database, you can drop those [databases](#) after upgrade completes.

- **Cloudera Management Service TLS/SSL configuration**

If you have enabled TLS security for the Cloudera Manager Admin Console, beginning with Cloudera Manager 5.1, Cloudera Management Service roles communicate with Cloudera Manager using TLS, and fail to start until TLS/SSL properties have been configured.

- **Impala**

Cloudera Manager 5 supports Impala 1.2.1 and higher. If the version of your Impala service is 1.1 or lower, the following upgrade instructions will work, but once the upgrade has completed, you will see a validation warning for your Impala service. You will not be able to restart your Impala (or Hue) services until you upgrade your Impala service to 1.2.1 and higher. If you want to continue to use Impala 1.1 and lower, *do not* upgrade to Cloudera Manager 5.

- **Navigator**

If you have enabled auditing with Cloudera Navigator, during the upgrade to Cloudera Manager 5, auditing is suspended and is only restarted when you restart the roles of audited services.

- **JDK upgrade**

If you upgrade the JDK during the installation of the Cloudera Manager Agent, you must *restart all services*. Additionally, if you have enabled TLS/SSL, you must reinstall CA certificates to your truststores. See [Creating and Using the Truststore](#).

- **Hard Restart of Cloudera Manager Agents**

Certain circumstances require that you hard restart the Cloudera Manager Agent on each host:

- Deploying a fix to an issue where Cloudera Manager did not always correctly restart services
- Using the maximum file descriptor feature
- Enabling HDFS DataNodes to start if you perform the step [\(Optional\) Upgrade CDH](#) on page 52 after upgrading Cloudera Manager



Important:

- **Hive**

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

When upgrading from Cloudera Manager versions lower than 4.5, Cloudera Manager automatically creates new Hive services to capture the previous implicit Hive dependency from Hue and Impala. Your previous services continue to function without impact. If Hue was using a Hive metastore backed by a Derby database, the newly created Hive Metastore Server also uses Derby. Because Derby does not allow concurrent connections, Hue continues to work, but the new Hive Metastore Server does not run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve cluster functionality as it existed before upgrade.

Cloudera recommends switching to a different supported database because of the limitations of a Derby-backed Hive metastore.

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. In releases lower than Cloudera Manager 4.5, Hue and Impala connected directly to the Hive metastore database, so the bypass mode is enabled by default when upgrading to Cloudera Manager 4.5 and higher. This ensures that the upgrade does not disrupt your existing setup. You should plan to disable the bypass mode, especially when using CDH 4.2 and higher. Using the Hive Metastore Server is the recommended configuration, and the WebHCat Server role requires the Hive Metastore Server to *not* be bypassed. To disable bypass mode, see [Disabling Bypass Mode](#).

Cloudera Manager 4.5 and higher also supports HiveServer2 with CDH 4.2. In CDH 4, HiveServer2 is not added by default, but can be added as a new role under the Hive service (see [Role Instances](#)). In CDH 5, HiveServer2 is a mandatory role.



Note: If you are upgrading from Cloudera Manager Free Edition 4.5 and lower, you are upgraded to Cloudera Express, which includes a number of features that were previously available only with Cloudera Enterprise. Of those features, activity monitoring requires a database. When upgrading to Cloudera Manager 5, you must specify Activity Monitor database information. You can use the embedded PostgreSQL database, which Cloudera Manager can set up automatically.

Before You Begin



Warning: Cloudera Manager 5 does not support CDH 3 and you cannot upgrade Cloudera Manager 4 to Cloudera Manager 5 if you have a cluster running CDH 3. To upgrade CDH 3 clusters to CDH 4 using Cloudera Manager, you must use Cloudera Manager 4.

Perform the following *before* upgrading to Cloudera Manager 5:

- **Obtain host credentials** - For Cloudera Manager to upgrade the Agent packages, you must have SSH access and be able to log in using a root account or an account that has password-less sudo permission. See [CDH and Cloudera Manager Networking and Security Requirements](#) for more information.
- **Prepare databases** - See [Database Considerations for Cloudera Manager Upgrades](#) on page 160.
- Cloudera Manager 5 supports HDFS high availability only with automatic failover. If your cluster has enabled high availability without automatic failover, you must enable automatic failover before upgrading to Cloudera Manager 5. See [Configuring HDFS High Availability](#).

Stop Selected Services and Roles

If your cluster meets any of the conditions listed in the following table, you must stop the indicated services or roles.

Condition	Procedure
Running a version of Cloudera Manager that has the Cloudera Management Service	Stop the Cloudera Management Service.
Upgrading from Cloudera Manager 4.5 and higher, and using the embedded PostgreSQL database for the Hive metastore	<p>Stop the services that have a dependency on the Hive metastore (Hue, Impala, and Hive). You cannot stop the Cloudera Manager Server database while these services are running. If you attempt to upgrade while the embedded database is running, the upgrade fails. Stop services that depend on the Hive metastore in the following order:</p> <ol style="list-style-type: none"> 1. Stop the Hue and Impala services. 2. Stop the Hive service.
Running Cloudera Navigator	<p>Stop any of the following roles for which the service's Queue Policy configuration (<code>navigator.batch.queue_policy</code>) is set to SHUTDOWN:</p> <ul style="list-style-type: none"> • HDFS - NameNode • HBase - Master and RegionServers • Hive - HiveServer2 • Hue - Beeswax Server <p>Stopping these roles makes any service depending on these roles unavailable. For HDFS - NameNode, most of the services in the cluster are unavailable until the upgrade is finished.</p>

Stop Cloudera Manager Server, Database, and Agent

1. Use the Admin Console to stop any running commands. These include user commands and commands Cloudera Manager automatically triggers in response to a state change or a schedule. You can either wait for commands to complete or stop any running commands. For more information on viewing and stopping running commands, see [Viewing Running and Recent Commands](#). If you do not stop all commands, the Cloudera Manager Server *fails to start after upgrade*.
2. On the host running the Cloudera Manager Server, stop the Cloudera Manager Server:

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

3. If you are using the embedded PostgreSQL database for Cloudera Manager, stop the database:

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



Important: If you are *not* running the embedded database service and you attempt to stop it, you receive a message indicating that the service cannot be found. If instead you get a message that the shutdown failed, the embedded database is still running, probably because services are connected to the Hive metastore. If the database shutdown fails due to connected services, issue the following command:

- RHEL-compatible 7 and higher:

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

- All other Linux distributions:

```
sudo service cloudera-scm-server-db fast_stop
```

4. If the Cloudera Manager host is also running the Cloudera Manager Agent, stop the Agent:

```
sudo service cloudera-scm-agent hard_stop_confirmed
```

(Optional) Upgrade the JDK on Cloudera Manager Server and Agent Hosts

If you are manually upgrading the Cloudera Manager Agent software in [Upgrade and Start Cloudera Manager Agents \(Packages\)](#) on page 42 or [Install Cloudera Manager Server and Agent Software \(Tarballs\)](#) on page 38, and you are upgrading to CDH 5, install the Oracle JDK on the Agent hosts as described in [Java Development Kit Installation](#).

If you are not running Cloudera Manager Server on the same host as a Cloudera Manager Agent, and you want all hosts to run the same JDK version, install the Oracle JDK on that host.

Upgrade Cloudera Manager Software

Choose a procedure based on how you installed Cloudera Manager.

Upgrade Cloudera Manager Server (Packages)

1. To upgrade the Cloudera Manager Server packages, you can upgrade from the Cloudera repository at <https://archive.cloudera.com/cm5/>, or you can create your own repository, as described in [Understanding Custom Installation Solutions](#). You must create your own repository if you are upgrading a cluster that does not have Internet access.
 - a. Find the Cloudera repo file for your distribution by starting at <https://archive.cloudera.com/cm5/> and navigating to the directory that matches your operating system.
 - For RHEL or CentOS 6, go to https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/. In that directory, find the repo file that contains information including the repository base URL and GPG key. The contents of `cloudera-manager.repo` are similar to the following:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 5, on RHEL or CentOS 6 x86_64
name=Cloudera Manager
baseurl=https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5/
gpgkey = https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/RPM-GPG-KEY-cloudera
gpgcheck = 1
```


- For Ubuntu or Debian systems, go to the appropriate release directory, for example, <https://archive.cloudera.com/cm4/debian/wheezy/amd64/cm>. The repo file, in this case `cloudera.list`, is similar to the following:

```
# Packages for Cloudera Manager, Version 5, on Debian 7.0 x86_64
deb https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
deb-src https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
```

- b. Replace the repo file in the configuration location for the package management software for your system.

Operating System	Commands
RHEL	Copy <code>cloudera-manager.repo</code> to <code>/etc/yum.repos.d/</code> .
SLES	Copy <code>cloudera-manager.repo</code> to <code>/etc/zypp/repos.d/</code> .
Ubuntu or Debian	Copy <code>cloudera.list</code> to <code>/etc/apt/sources.list.d/</code> .

- c. Run the following commands:

Operating System	Commands
RHEL	<pre>\$ sudo yum clean all \$ sudo yum upgrade cloudera-manager-server cloudera-manager-daemons cloudera-manager-agent</pre> <div>  Note: <ul style="list-style-type: none"> <code>yum clean all</code> cleans <code>yum</code> cache directories, ensuring that you download and install the latest versions of the packages. If your system is not up to date, any underlying system components must be upgraded before <code>yum update</code> can complete. <code>yum</code> indicates which components must be upgraded. If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add <code>cloudera-manager-server-db-2</code> to the list of packages in the <code>yum upgrade</code> command. <i>The embedded PostgreSQL database should not be used in production environments.</i> </div>
SLES	<pre>\$ sudo zypper clean --all \$ sudo zypper up -r https://archive.cloudera.com/cm5/sles/11/x86_64/cm/5/</pre> <p>To download from your own repository:</p> <pre>\$ 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</pre>
Ubuntu or Debian	<p>The following commands clean cached repository information and update Cloudera Manager components:</p> <pre>\$ sudo apt-get clean \$ sudo apt-get update \$ sudo apt-get dist-upgrade \$ sudo apt-get install cloudera-manager-server cloudera-manager-daemons cloudera-manager-agent</pre>

Operating System	Commands
	<p>If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add <code>cloudera-manager-server-db-2</code> to the list of packages in the <code>apt-get install</code> command. <i>The embedded PostgreSQL database should not be used in production environments.</i></p> <p>During this process, you may be prompted about your configuration file version:</p> <pre>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.</pre> <p>You will receive a similar prompt for <code>/etc/cloudera-scm-server/db.properties</code>. Answer N to both prompts.</p>

- If you customized `/etc/cloudera-scm-agent/config.ini`, your customized file is renamed with the extension `.rpmsave` or `.dpkg-old`. Merge any customizations into `/etc/cloudera-scm-agent/config.ini` installed by the package manager.

You should now have the following packages, corresponding to the version of Cloudera Manager you installed, on the host that will be the Cloudera Manager Server host.

OS	Packages
RPM-based distributions	<pre>\$ rpm -qa 'cloudera-manager-*' cloudera-manager-repository-5.0-1.noarch cloudera-manager-server-5.11.0-0.cm5110.p0.97.el6.x86_64 cloudera-manager-agent-5.11.0-0.cm5110.p0.97.el6.x86_64 cloudera-manager-daemons-5.11.0-0.cm5110.p0.97.el6.x86_64</pre>
Ubuntu or Debian	<pre>~# 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-agent 5.11.0-0.cm5110.p0.97~sq The Cloudera Manager Agent ii cloudera-manager-daemo 5.11.0-0.cm5110.p0.97~sq Provides daemons for monitoring Hadoop and related tools. ii cloudera-manager-serve 5.11.0-0.cm5110.p0.97~sq The Cloudera Manager Server</pre>

You may also see an entry for the `cloudera-manager-server-db-2` if you are using the embedded database, and additional packages for plug-ins, depending on what was previously installed on the server host. If the `cloudera-manager-server-db-2` package is installed, and you do not plan to use the embedded database, you can remove this package.

Install Cloudera Manager Server and Agent Software (Tarballs)

To install the Cloudera Manager Server and Agents, you download and extract tarballs, create users, and configure the server and agents.

Download and Extract Tarballs

Tarballs contain the Cloudera Manager Server and Cloudera Manager Agent in a single file.

1. Download tarballs from the locations listed in .
2. Copy the tarballs and unpack them on all hosts on which you intend to install Cloudera Manager Server and Cloudera Manager Agents, in a directory you choose. You can create a new directory to accommodate the files you extract from the tarball. For example, if `/opt/cloudera-manager` does not exist, create it using a command similar to:

```
$ sudo mkdir /opt/cloudera-manager
```

3. Extract the contents of the tarball to the selected directory. For example, to copy a tar file to your home directory and extract the contents of all tar files to the `/opt/` directory, use a command similar to the following:

```
$ sudo tar xzf cloudera-manager*.tar.gz -C /opt/cloudera-manager
```

The files are extracted to a subdirectory named according to the Cloudera Manager version being extracted. For example, files could be extracted to `/opt/cloudera-manager/cm-5.0/`. This full path is required later and is referred to as `$CMF_DEFAULTS` directory.

Perform Configuration Required by Single User Mode

If you are creating a Cloudera Manager deployment that employs single user mode, perform the configuration steps described in [Configuring Single User Mode](#).

Create Users

The Cloudera Manager Server and managed services require a user account. When installing Cloudera Manager from tarballs, you must create this user account on all hosts manually. Because Cloudera Manager Server and managed services are configured to use `tcoudera-scm` by default, creating a user with this name is the simplest approach. This created user is used automatically after installation is complete.

To create user `cloudera-scm`, use a command such as the following:

```
$ sudo useradd --system --home=/opt/cloudera-manager/cm-5.6.0/run/cloudera-scm-server
--no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm
```

Ensure that the `--home` argument path matches your environment. This argument varies according to where you place the tarball, and the version number varies among releases. For example, the `--home` location could be `/opt/cm-5.6.0/run/cloudera-scm-server`.

Create the Cloudera Manager Server Local Data Storage Directory

1. Create the following directory: `/var/lib/cloudera-scm-server`.
2. Change the owner of the directory so that the `cloudera-scm` user and group have ownership of the directory. For example:

```
$ sudo mkdir /var/lib/cloudera-scm-server
$ sudo chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-server
```

Configure Cloudera Manager Agents

On every Cloudera Manager Agent host, configure the Cloudera Manager Agent to point to the Cloudera Manager Server. Use the `CMF_DEFAULTS` environment variable in the environment of the user running Cloudera Manager Server and Agent. In `$CMF_DEFAULTS/cloudera-scm-agent/config.ini`, set the following environment variables:

Property	Description
server_host	Name of the host where Cloudera Manager Server is running.
server_port	Port on the host where Cloudera Manager Server is running.

By default, a tarball installation has a `var` subdirectory where state is stored. In a non-tarball installation, state is stored in `/var`. Cloudera recommends that you reconfigure the tarball installation to use an external directory as the `/var` equivalent (`/var` or any other directory outside the tarball) so that when you upgrade Cloudera Manager, the new tarball installation can access this state. Configure the installation to use an external directory for storing state by editing `$CMF_DEFAULTS/etc/default/cloudera-scm-agent` and setting the `CMF_VAR` variable to the location of the `/var` equivalent. If you do not reuse the state directory between different tarball installations, duplicate Cloudera Manager Agent entries can occur in the Cloudera Manager database.

Configuring for a Custom Cloudera Manager User and Custom Directories

You can change the default username and directories used by Cloudera Manager. If you do not change the default, skip to [Cloudera Manager and Managed Service Datastores](#). By default, Cloudera Manager creates the following directories in `/var/log` and `/var/lib`:

- `/var/log/cloudera-scm-headlamp`
- `/var/log/cloudera-scm-firehose`
- `/var/log/cloudera-scm-alertpublisher`
- `/var/log/cloudera-scm-eventserver`
- `/var/lib/cloudera-scm-headlamp`
- `/var/lib/cloudera-scm-firehose`
- `/var/lib/cloudera-scm-alertpublisher`
- `/var/lib/cloudera-scm-eventserver`
- `/var/lib/cloudera-scm-server`

If you are using a custom username and custom directories for Cloudera Manager, you must create these directories on the Cloudera Manager Server host and assign ownership of these directories to the custom username. Cloudera Manager installer makes no changes to any directories that already exist. Cloudera Manager cannot write to any existing directories for which it does not have proper permissions, and if you do not change ownership, Cloudera Management Service roles may not perform as expected. To resolve these issues, do one of the following:

- **Change ownership of existing directories:**

Use the `chown` command to change ownership of all existing directories to the Cloudera Manager user. If the Cloudera Manager username and group are `cloudera-scm`, to change the ownership of the headlamp log directory, issue a command similar to the following:

```
$ sudo chown -R cloudera-scm:cloudera-scm /var/log/cloudera-scm-headlamp
```

- **Use alternate directories:**

1. If the directories you plan to use do not exist, create them. For example, to create `/var/cm_logs/cloudera-scm-headlamp` for use by the `cloudera-scm` user, run the following commands:

```
mkdir /var/cm_logs/cloudera-scm-headlamp
chown cloudera-scm /var/cm_logs/cloudera-scm-headlamp
```

2. Connect to the Cloudera Manager Admin Console.
3. Select **Clusters > Cloudera Management Service**
4. Select **Scope > role name**.
5. Click the **Configuration** tab.
6. Enter a term in the **Search** field to find the settings to be changed. For example, you can enter `/var` or `directory`.

7. Update each value with the new locations for Cloudera Manager to use.



Note: The configuration property for the **Cloudera Manager Server Local Data Storage Directory** (/var/lib/cloudera-scm-server by default) is located on a different page:

1. Select **Administration > Settings**.
2. Type `directory` in the Search box.
3. Enter the directory path in the **Cloudera Manager Server Local Data Storage Directory** property.

8. Click **Save Changes** to commit the changes.

Start Cloudera Manager Server

Choose a procedure based on how you installed Cloudera Manager.

Start the Cloudera Manager Server (Packages)

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

1. If you are using the embedded PostgreSQL database for Cloudera Manager, start the database:

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

2. Start the Cloudera Manager Server:

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

You should see the following:

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

If the Cloudera Manager Server does not start, see [Troubleshooting Installation and Upgrade Problems](#).

Start the Cloudera Manager Server (Tarballs)

The way in which you start the Cloudera Manager Server varies according to which account you want the Server to run under:

- As root:

```
$ sudo $CMF_DEFAULTS/etc/init.d/cloudera-scm-server start
```

- As another user. If you run as another user, ensure that the user you created for Cloudera Manager owns the location to which you extracted the tarball including the newly created database files. If you followed the earlier examples and created the directory `/opt/cloudera-manager` and the user `cloudera-scm`, you could use the following command to change ownership of the directory:

```
$ sudo chown -R cloudera-scm:cloudera-scm /opt/cloudera-manager
```

Once you have established ownership of directory locations, you can start Cloudera Manager Server using the user account you chose. For example, you might run the Cloudera Manager Server as `cloudera-service`. In this case, you have the following options:

- Run the following command:

```
$ sudo -u cloudera-service $CMF_DEFAULTS/etc/init.d/cloudera-scm-server start
```

Upgrading Cloudera Manager

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from `$CMF_DEFAULTS/etc/default/cloudera-scm-server`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `$CMF_DEFAULTS/etc/init.d/cloudera-scm-server` to the user you want the server to run as. For example, to run as `cloudera-service`, change the user and group as follows:

```
USER=cloudera-service  
GROUP=cloudera-service
```

3. Run the server script as root:

```
$ sudo $CMF_DEFAULTS/etc/init.d/cloudera-scm-server start
```

- To start the Cloudera Manager Server automatically after a reboot:

1. Run the following commands on the Cloudera Manager Server host:

- **RHEL-compatible and SLES**

```
$ cp $CMF_DEFAULTS/etc/init.d/cloudera-scm-server /etc/init.d/cloudera-scm-server  
$ chkconfig cloudera-scm-server on
```

- **Debian/Ubuntu**

```
$ cp $CMF_DEFAULTS/etc/init.d/cloudera-scm-server /etc/init.d/cloudera-scm-server  
$ update-rc.d cloudera-scm-server defaults
```

2. On the Cloudera Manager Server host, open the `/etc/init.d/cloudera-scm-server` file and change the value of `CMF_DEFAULTS` from `${CMF_DEFAULTS:-/etc/default}` to `$CMF_DEFAULTS/etc/default`.

Upgrade and Start Cloudera Manager Agents

Choose a procedure based on how you installed Cloudera Manager.

Upgrade and Start Cloudera Manager Agents (Packages)



Important: All hosts in the cluster must have access to the Internet if you 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.
2. Upgrade hosts using one of the following methods:
 - **Cloudera Manager installs Agent software**
 1. Select **Yes, I would like to upgrade the Cloudera Manager Agent packages now** and click **Continue**.
 2. Select the release of the Cloudera Manager Agent to install. Normally, this is the **Matched Release for this Cloudera Manager Server**. However, if you used a custom repository (instead of `archive.cloudera.com`) for the Cloudera Manager server, select **Custom Repository** and provide the required information. The custom repository allows you to use an alternative location, but that location must contain the matched Agent version.
 3. Click **Continue**. The JDK Installation Options page displays.

- Leave **Install Oracle Java SE Development Kit (JDK)** checked to allow Cloudera Manager to install the JDK on each cluster host, or uncheck if you plan to install it yourself.
- If local laws permit you to deploy unlimited strength encryption, and you are running a secure cluster, check the **Install Java Unlimited Strength Encryption Policy Files** checkbox.

Click **Continue**.

4. Specify credentials and initiate Agent installation:

- Select **root** or enter the username for an account that has password-less sudo permission.
- Select an authentication method:
 - If you choose password authentication, enter and confirm the password.
 - If you choose public-key authentication, provide a passphrase and path to the required key files.
- You can 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.

5. Click **Continue**. The Cloudera Manager Agent packages are installed.


6. Click **Continue**. The Host Inspector runs to inspect your managed hosts for correct versions and configurations. If there are problems, you can make changes and then rerun the inspector. When you are satisfied with the inspection results, click **Finish**.

- **Manually install Agent software**

1. On all cluster hosts except the Cloudera Manager Server host, stop the Agent:

```
sudo service cloudera-scm-agent hard_stop_confirmed
```

2. In the Cloudera Admin Console, select **No, I would like to skip the agent upgrade now** and click **Continue**.
3. Copy the appropriate repo file as described in [Upgrade Cloudera Manager Server \(Packages\)](#) on page 36.
4. Run the following commands:

Operating System	Commands
RHEL	<pre>\$ sudo yum clean all \$ sudo yum upgrade cloudera-manager-server cloudera-manager-daemons cloudera-manager-agent</pre> <div>  Note: <ul style="list-style-type: none"> • <code>yum clean all</code> cleans <code>yum</code> cache directories, ensuring that you download and install the latest versions of the packages. • If your system is not up to date, any underlying system components must be upgraded before <code>yum update</code> can complete. <code>yum</code> indicates which components must be upgraded. • If the Cloudera Manager instance you are upgrading uses the embedded PostgreSQL database, add <code>cloudera-manager-server-db-2</code> to the list of packages in the <code>yum upgrade</code> command. <i>The embedded PostgreSQL database should not be used in production environments.</i> </div>
SLES	<pre>\$ sudo zypper clean --all \$ sudo zypper up -r https://archive.cloudera.com/cm5/sles/11/x86_64/cm/5/</pre> <p>To download from your own repository:</p> <pre>\$ 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</pre>
Ubuntu or Debian	<p>Use the following commands to clean cached repository information and update Cloudera Manager components:</p> <pre>\$ sudo apt-get clean \$ sudo apt-get update \$ sudo apt-get dist-upgrade \$ sudo apt-get install cloudera-manager-agent cloudera-manager-daemons</pre> <p>During this process, you may be prompted about your configuration file version:</p> <pre>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.</pre> <p>You will receive a similar prompt for <code>/etc/cloudera-scm-server/db.properties</code>. Answer N to both prompts.</p>

5. If you customized `/etc/cloudera-scm-agent/config.ini`, your customized file is renamed with the extension `.rpmsave` or `.dpkg-old`. Merge any customizations into `/etc/cloudera-scm-agent/config.ini` installed by the package manager.

6. On all cluster hosts, start the Agent:

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

3. If you are upgrading from a free version of Cloudera Manager lower than 4.6:

- a. Click **Continue** to assign the Cloudera Management Services roles to hosts.
- b. If you are upgrading to Cloudera Enterprise, specify required databases:
 - a. Choose the database type:
 - Keep the default setting of **Use Embedded Database** to have Cloudera Manager create and configure required databases. Record the auto-generated passwords.

Cluster Setup

Database Setup

Configure and test database connections. If using custom databases, create the databases first according to the [Installing and Configuring an External Database](#) section of the [Installation Guide](#).

☐ Use Custom Databases
☒ Use Embedded Database

When using the embedded database, passwords are automatically generated. Please copy them down.

Service	Database Host Name	Database Type	Database Name	Username	Password	Status
Hive	tcdn2-1.ent.cloudera.com:7432	PostgreSQL	hive	hive	t56lwbdk4F	✓ Skipped. Cloudera Manager will create this database in a later step.
Reports Manager	tcdn2-1.ent.cloudera.com:7432	PostgreSQL	rman	rman	Y6S4lWvfNo	✓ Successful
Navigator Audit Server	tcdn2-1.ent.cloudera.com:7432	PostgreSQL	nav	nav	QLR2B0qqO9	✓ Successful
Navigator Metadata Server	tcdn2-1.ent.cloudera.com:7432	PostgreSQL	navms	navms	lmo07jxOen	✓ Successful
Oozie Server	tcdn2-1.ent.cloudera.com:7432	PostgreSQL	oozie_oozie_se	oozie_oozie_se	NTF1KNdpPI	✓ Skipped. Cloudera Manager will create this database in a later step.

[Test Connection](#)

- Select **Use Custom Databases** to specify the external database host and enter the database type, database name, username, and password for the custom database.
 - If you are adding the Oozie service, you can change your Oozie configuration to control when data is purged to improve performance, reduce database disk usage, improve upgrade performance, or to keep the history for a longer period of time. See [Configuring Oozie Data Purge Settings Using Cloudera Manager](#).
- b. Click **Test Connection** to confirm that Cloudera Manager can communicate with the database using the information you have supplied. If the test succeeds in all cases, click **Continue**; otherwise, check and correct the information you have provided for the database and then try the test again. (For some servers, if you are using the embedded database, you will see a message saying the database will be created at a later step in the installation process.)

The **Cluster Setup Review Changes** screen displays.

4. Click **Finish**.

5. If you are upgrading from Cloudera Manager prior to 4.5:

- a. Select the host for the Hive Metastore Server role.
- b. Review the configuration values and click **Accept** to continue.



Note:

- If Hue was using a Hive metastore backed by a Derby database, the newly created Hive Metastore Server also uses Derby. Because Derby does not allow concurrent connections, Hue continues to work, but the new Hive Metastore Server does not run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve cluster functionality as it existed before upgrade. Cloudera recommends switching to a different supported database because of the limitations of a Derby-backed Hive metastore.
- In releases lower than Cloudera Manager 4.5, Hue and Impala connected directly to the Hive metastore database, so the bypass mode is enabled by default when upgrading to Cloudera Manager 4.5 and higher. This ensures that the upgrade does not disrupt your existing setup. You should plan to disable the bypass mode, especially when using CDH 4.2 and higher. Using the Hive Metastore Server is the recommended configuration, and the WebHCat Server role requires the Hive Metastore Server to *not* be bypassed. To disable bypass mode, see [Disabling Bypass Mode](#). After changing this configuration, you must redeploy your client configurations, restart Hive, and restart any Hue or Impala services configured to use that Hive.
- If you are using CDH 4.0 or CDH 4.1, see known issues related to Hive in [Known Issues and Workarounds in Cloudera Manager 5](#).

6. If you are upgrading from Cloudera Manager 4.5 or lower, correct the Hive home directory permissions (/user/hive) as follows:

```
sudo -u hdfs hdfs dfs -chown hive:hive /user/hive/  
sudo -u hdfs hdfs dfs -chmod 1775 /user/hive/
```

7. If you are upgrading from Cloudera Manager 4.7 or lower and have an Impala service, assign the Impala Catalog Server role to a host.
8. Review the configuration changes to be applied.
9. Click **Finish**.

All services (except for the services you stopped in [Stop Selected Services and Roles](#) on page 35) should be running.

Restart Cloudera Manager Agents (Tarballs)

Stop Cloudera Manager Agents (Tarballs)

- To stop the Cloudera Manager Agent, run this command on each Agent host:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent stop
```

- If you are running [single user mode](#), stop Cloudera Manager Agent using the user account you chose. For example, if you are running the Cloudera Manager Agent as `cloudera-scm`, you have the following options:

- Run the following command:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent stop
```

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from `tarball_root/etc/default/cloudera-scm-agent`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `tarball_root/etc/init.d/cloudera-scm-agent` to the user you want the Agent to run as. For example, to run as `cloudera-scm`, change the user and group as follows:

```
USER=cloudera-scm
GROUP=cloudera-scm
```

3. Run the Agent script as root:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent stop
```

Start Cloudera Manager Agents (Tarballs)

Start the Cloudera Manager Agent according to the account you want the Agent to run under:

- To start the Cloudera Manager Agent, run this command on each Agent host:

```
$ sudo $CMF_DEFAULTS/etc/init.d/cloudera-scm-agent start
```

When the Agent starts, it contacts the Cloudera Manager Server.

- If you are running [single user mode](#), start Cloudera Manager Agent using the user account you chose. For example, to run the Cloudera Manager Agent as `cloudera-scm`, you have the following options:
 - Run the following command:

```
$ sudo -u cloudera-scm $CMF_DEFAULTS/etc/init.d/cloudera-scm-agent start
```

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from `$CMF_DEFAULTS/etc/default/cloudera-scm-agent`:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in `$CMF_DEFAULTS/etc/init.d/cloudera-scm-agent` to the user you want the Agent to run as. For example, to run as `cloudera-scm`, change the user and group as follows:

```
USER=cloudera-scm
GROUP=cloudera-scm
```

3. Run the Agent script as root:

```
$ sudo $CMF_DEFAULTS/etc/init.d/cloudera-scm-agent start
```

- To start the Cloudera Manager Agents automatically after a reboot:

1. Run the following commands on each Agent host:

- **RHEL-compatible and SLES**

```
$ cp $CMF_DEFAULTS/etc/init.d/cloudera-scm-agent /etc/init.d/cloudera-scm-agent
$ chkconfig cloudera-scm-agent on
```

- **Debian/Ubuntu**

```
$ cp $CMF_DEFAULTS/etc/init.d/cloudera-scm-agent /etc/init.d/cloudera-scm-agent
$ update-rc.d cloudera-scm-agent defaults
```

2. On each Agent, open the `$CMF_DEFAULTS/etc/init.d/cloudera-scm-agent` file and change the value of `CMF_DEFAULTS` from `${CMF_DEFAULTS:-/etc/default}` to `$CMF_DEFAULTS/etc/default`.

Verify that the Upgrade Succeeded

If the commands to update and start the Cloudera Manager Server complete without errors, you can assume the upgrade has completed successfully. To verify, check that the server versions have been updated.

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.

Add Hive Gateway Roles

If you are upgrading from Cloudera Manager 4.4 or lower, add Hive gateway roles to any hosts where Hive clients should run.

1. Go to the Hive service.
2. Click the **Instances** tab.
3. Click the **Add Role Instances** button.
4. Select the hosts on which you want a Hive gateway role to run. This ensures that the Hive client configurations are deployed on these hosts.

Configure the Cluster Version for Package Installs

If you have installed CDH as a package, after an installation or upgrade, make sure that the cluster CDH version matches the package CDH version, using the procedure in [Configuring the CDH Version of a Cluster](#) on page 113. If the cluster CDH version does not match the package CDH version, Cloudera Manager incorrectly enables and disables service features based on the cluster's configured CDH version.

Upgrade Impala

If your version of Impala is 1.1 and lower, upgrade to Impala 1.2.1 and higher.

(Optional) Configure TLS/SSL for Cloudera Management Service

If you have enabled TLS security for the Cloudera Manager Admin Console, beginning with Cloudera Manager 5.1, Cloudera Management Service roles communicate with Cloudera Manager using TLS, and fail to start until TLS/SSL properties have been configured. Configure Cloudera Management Service roles to communicate with Cloudera Manager over TLS/SSL as follows:

1. Do one of the following:
 - Select **Clusters > Cloudera Management Service**.
 - On the **Home > Status** tab, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.
2. Click the **Configuration** tab.
3. Select **Scope > Cloudera Management Service (Service-Wide)**.
4. Select **Category > Security**.
5. Edit the following TLS/SSL properties according to your cluster configuration.

Property	Description
TLS/SSL Client Truststore File Location	Path to the client truststore file used in HTTPS communication. The contents of this truststore can be modified without restarting the Cloudera Management Service roles. By default, changes to its contents are picked up within ten seconds.

Property	Description
TLS/SSL Client Truststore File Password	Password for the client truststore file.

6. Click **Save Changes** to commit the changes.

7. Restart the Cloudera Management Service.

For more information, see [Configuring Cloudera Manager Clusters for TLS/SSL](#).

(Optional) Deploy a Cloudera Manager Agent Upgrade

Several conditions require you to upgrade the Cloudera Manager Agent:

- Deploying a fix to an issue where Cloudera Manager did not always correctly restart services
- Using the maximum file descriptor feature
- Enabling HDFS DataNodes to start if you perform the step [\(Optional\) Upgrade CDH](#) on page 52 after upgrading Cloudera Manager

To deploy the Cloudera Manager Agent upgrade, perform the following steps:

1. Stop all CDH and managed services.
2. On all hosts with Cloudera Manager Agents, hard-restart the Agents. Before performing this step, ensure you understand the semantics of the `hard_restart` command by reading [Hard Stopping and Restarting Agents](#).

- Packages

- RHEL-compatible 7 and higher:

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

- All other Linux distributions:

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

- Tarballs

- To stop the Cloudera Manager Agent, run this command on each Agent host:

- RHEL-compatible 7 and higher:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent next_stop_hard
$ sudo tarball_root/etc/init.d/cloudera-scm-agent restart
```

- All other Linux distributions:

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent hard_restart
```

- If you are running [single user mode](#), start Cloudera Manager Agent using the user account you chose. For example, to run the Cloudera Manager Agent as `cloudera-scm`, you have the following options:

- Run the following command:

- RHEL-compatible 7 and higher:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent next_stop_hard
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent restart
```

- All other Linux distributions:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent hard_restart
```

- Edit the configuration files so the script internally changes the user, and then run the script as root:

1. Remove the following line from *tarball_root/etc/default/cloudera-scm-agent*:

```
export CMF_SUDO_CMD=" "
```

2. Change the user and group in *tarball_root/etc/init.d/cloudera-scm-agent* to the user you want the Agent to run as. For example, to run as *cloudera-scm*, change the user and group as follows:

```
USER=cloudera-scm  
GROUP=cloudera-scm
```

3. Run the Agent script as root:

- RHEL-compatible 7 and higher:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent next_stop_hard  
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent restart
```

- All other Linux distributions:

```
$ sudo -u cloudera-scm tarball_root/etc/init.d/cloudera-scm-agent hard_restart
```

3. Start all services.

Deploy a JDK Upgrade

If you upgraded the JDK when installing the Cloudera Manager Agents, do the following:

1. If the Cloudera Manager Server host is also running a Cloudera Manager Agent, restart the Cloudera Manager Server:

- Packages

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

- Tarballs

```
$ sudo tarball_root/etc/init.d/cloudera-scm-agent restart
```

If the Cloudera Manager Server does not start, see [Troubleshooting Installation and Upgrade Problems](#).

2. If you have not restarted services in previous steps, restart all services:

- a. On the **Home > Status** tab, click



next to the cluster name and select **Restart**.

- b. In the confirmation dialog box, click **Restart**.

(Optional) Deploy a Monitoring Upgrade

Cloudera Manager 5 provides monitoring support for all roles. However, the Cloudera Manager Agent does not send monitoring data for these roles until:

1. The Cloudera Manager Agent has been upgraded and restarted.
2. The monitored roles have been restarted.

Until you restart the roles, some data is not in the monitoring charts and health tests. To enable monitoring for all roles, if you have not restarted services in previous steps, perform the following steps:

1. On the **Home > Status** tab, click



next to the cluster name and select **Restart**.

2. In the confirmation dialog box, click **Restart**.

Start Selected Services and Roles

Start services and roles you shut down in [Stop Selected Services and Roles](#) on page 35 that have not been started in other steps:

1. If you are not upgrading CDH, do the following:

- a. If you are running Cloudera Navigator, start the following roles of audited services:

- **HDFS** - NameNode
- **HBase** - Master and RegionServers
- **Hive** - HiveServer2
- **Hue** - Beeswax Server

- b. On the **Home > Status** tab, click



next to the name of each service you shut down and select **Start**.

- c. In the confirmation dialog box, click **Start**.


2. On the **Home > Status** tab, click



next to the Cloudera Management Service and select **Start**.

3. In the confirmation dialog box, click **Start**.

Deploy Updated Client Configurations

The services for which client configurations require redeployment are indicated with the  icon on the **Home > Status** tab. To ensure that clients have current information about resources, update the client configuration:

1. On the **Home > Status** tab, click



next to the cluster name and select **Deploy Client Configuration**.

2. In the confirmation dialog box, click **Deploy Client Configuration**.

Test the Installation

When you have finished the upgrade to Cloudera Manager, test the installation to verify that the monitoring features are working as expected; follow the instructions in [Testing the Installation](#).

(Optional) Upgrade CDH

Cloudera Manager 5 can manage both CDH 4 and CDH 5, so upgrading existing CDH 4 installations is not required, but you may want to upgrade to the latest version. For more information on upgrading CDH, see [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54.

Package Dependencies

When upgrading Cloudera Manager note that the following packages may be installed or upgraded as part of the steps taken by the upgrade wizard.

RHEL/CentOS

- bind-utils
- chkconfig
- cyrus-sasl-gssapi
- cyrus-sasl-plain
- fuse
- fuse-libs
- gcc
- httpd
- init-functions
- libxslt
- mod_ssl
- MySQL-python
- openssl
- openssl-devel
- openssl-devel
- perl
- portmap
- postgresql-server >= 8.4
- psmisc
- python >= 2.4.3-43
- python-devel >= 2.4.3-43
- python-psycopg2
- python-setuptools
- sed
- service
- sqlite
- swig
- useradd
- zlib

SLES

- apache2
- bind-utils
- chkconfig
- cyrus-sasl-gssapi
- cyrus-sasl-plain
- fuse
- gcc
- libfuse2

- libxslt
- openssl
- openssl-devel
- perl
- portmap
- postgresql-server >= 8.4
- psmisc
- python >= 2.4.3-43
- python-devel >= 2.4.3-43
- python-mysql
- python-setuptools
- python-xml
- sed
- service
- sqlite
- swig
- useradd
- zlib

Debian/Ubuntu

- ant
- apache2
- bash
- chkconfig
- debhelper (>= 7)
- fuse-utils | fuse
- gcc
- libfuse2
- libssl2-modules
- libssl2-modules-gssapi-mit
- libsqlite3-0
- libssl-dev
- libxslt1.1
- lsb-base
- make
- openssl
- perl
- postgresql-client@@PG_PKG_VERSION@@
- postgresql@@PG_PKG_VERSION@@
- psmisc
- python-dev (>=2.4)
- python-mysqldb
- python-psycopg2
- python-setuptools
- rpcbind
- sed
- swig
- useradd
- zlib1g

Upgrading CDH and Managed Services Using Cloudera Manager

You can use Cloudera Manager to upgrade CDH for [major, minor, and maintenance upgrades](#). The procedures vary depending on the version of Cloudera Manager you are using and from which versions of CDH you are upgrading. Procedures to upgrade Cloudera Manager installations are different when using parcels compared to packages.

After completing preparatory steps, you use the Cloudera Manager upgrade wizard to complete the upgrade. If you use parcels (recommended), have enabled [HDFS High Availability](#), and have a Cloudera Enterprise license, you can perform a *rolling upgrade* that does not require you to take the cluster offline during the upgrade.

The Cloudera Manager minor version must always be *equal to or greater than* the CDH minor version because older versions of Cloudera Manager may not support features in newer versions of CDH. For example, if you want to upgrade to CDH 5.4.8, you must first upgrade to Cloudera Manager 5.4 or higher. To upgrade Cloudera Manager, see [Overview of Upgrading Cloudera Manager](#) on page 9.

Choose one of the following procedures to upgrade CDH using Cloudera Manager:

Upgrading CDH 5.x

The procedures for upgrading to CDH 5.x differ for minor release upgrades and maintenance release upgrades.

Minor Version Upgrades

To upgrade CDH 5.x to a higher version of CDH, for example from CDH 5.4 to CDH 5.8, select one of the following options:

- [Upgrading to CDH 5.x Using a Rolling Upgrade](#) on page 55

This option uses Cloudera Manager [parcels](#) and allows you to upgrade your cluster without having to stop the cluster. You must have HDFS high availability enabled and have a Cloudera Enterprise license.

- [Upgrading to CDH 5.x Using Parcels](#) on page 62

This option upgrades your cluster using parcels, but requires you to restart the cluster before completing the upgrade.

- [Upgrading to CDH 5.x Using Packages](#) on page 68

This option is the most time consuming and requires you to log in using `ssh` and execute a series of package commands on *all hosts* in your cluster. Cloudera recommends that you instead upgrade your cluster using [parcels](#), which allows Cloudera Manager to distribute the upgraded software to all hosts in the cluster without having to log in to each host. If you installed the cluster using packages, you can upgrade using parcels and the cluster will use parcels for subsequent upgrades.

Maintenance Version Upgrades

Maintenance upgrades fix critical bugs or address security issues. No new functionality or incompatible changes are introduced. The version numbers for maintenance releases differ only in the third digit, for example, when upgrading from CDH 5.8.0 to CDH 5.8.2.

To upgrade to a maintenance release, you only need to perform a subset of the Minor version upgrade steps. Follow the same procedures as for minor version upgrades but skip the steps that are labeled as follows:

[Not required for CDH maintenance release upgrades.]

Upgrading CDH 4 to CDH 5

You can upgrade CDH 4.x to CDH 5.x using either parcels or packages. These upgrades require you to take the cluster offline during the upgrade; rolling upgrades are not possible for this type of upgrade.

See [Upgrading CDH 4 to CDH 5](#) on page 83.

You can roll back an upgrade from CDH 4 to CDH 5, subject to some limitations. See [Rolling Back a CDH 4-to-CDH 5 Upgrade](#) on page 117.

Upgrading CDH 4.x

You can upgrade CDH 4.x to a higher minor release of CDH 4.x using either packages or parcels. You can also perform a rolling upgrade that does not require cluster down time.



Important: Cloudera Manager 4 and CDH 4 reached End of Maintenance (EOM) on August 9, 2015. Cloudera does not support or provide updates for Cloudera Manager 4 and CDH 4 releases. Cloudera recommends that you upgrade to Cloudera Manager 5 and CDH 5.

See:

- [Upgrading CDH 4](#) on page 104
- [Performing a Rolling Upgrade on a CDH 4 Cluster](#) on page 113

Upgrading CDH 3

To upgrade CDH 3 to CDH 4 using Cloudera Manager, follow the instructions in the CDH 4 documentation: [Upgrading CDH 3 to CDH 4 in a Cloudera Manager Deployment](#).



Important: Cloudera Manager 3 and CDH 3 reached End of Maintenance (EOM) on June 20, 2013. Cloudera does not support or provide updates for Cloudera Manager 3 and CDH 3 releases.

Upgrading to CDH 5.x Using a Rolling Upgrade

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

This topic describes how to perform a rolling upgrade from any version of CDH 5.x to a higher version of CDH 5.x, using Cloudera Manager and parcels. The minor version of Cloudera Manager you use to perform the upgrade must be equal to or greater than the CDH minor version. To upgrade Cloudera Manager, see [Overview of Upgrading Cloudera Manager](#) on page 9.

A rolling upgrade allows you to upgrade your cluster software and restart the upgraded services without taking the cluster offline. Performing a rolling upgrade requires the following:

- CDH was installed using Cloudera Manager and parcels. You can [migrate your cluster from using packages to using parcels](#).
- The cluster uses a Cloudera Enterprise license.
- High availability is enabled for HDFS.

**Note:**

This topic describes how to perform a rolling upgrade between maintenance and minor versions of CDH 5. For rolling upgrade between CDH 4 versions, see [Performing a Rolling Upgrade on a CDH 4 Cluster](#) on page 113.

You cannot perform a rolling upgrade from CDH 4 to CDH 5 because of incompatibilities between the two major versions. Instead, follow the instructions for a full upgrade at [Upgrading from CDH 4 to CDH 5 Parcels](#) on page 87.



Note: If you are upgrading to a *maintenance* version of CDH, skip any steps that are labeled **[Not required for CDH maintenance release upgrades.]**.

The version numbers for maintenance releases differ only in the third digit, for example when upgrading from CDH 5.8.0 to CDH 5.8.2. See [Maintenance Version Upgrades](#) on page 54.

To upgrade CDH using a rolling upgrade:

Step 1: Collect Upgrade Information

Before starting an upgrade, collect the following information:

1. Host credentials. You must have SSH access and be able to log in using a root account or an account that has password-less sudo permission.
2. The version of Cloudera Manager used in your cluster. Go to **Support > About**.
3. The version of the JDK deployed in the cluster. Go to **Support > About**.
4. The version of CDH. The CDH version number displays next to the cluster name on the **Home** page.
5. Whether the cluster was installed using parcels or packages. This information displays next to the CDH version on the **Home** page of Cloudera Manager.
6. The services enabled in your cluster. Go to **Clusters > Cluster name**.
7. Operating system type and version. Go to **Hosts** and click on a hostname in the list. The operating system type and version displays in the **Distribution** row in the **Details** section.
8. Database information for the databases used by Sqoop, Oozie, Hue, Hive Metastore, and Sentry Server (information is only required if these services are enabled in the cluster).

Gather the following information:

- Type of database (PostgreSQL, Embedded PostgreSQL, MySQL, MariaDB, or Oracle)
- Hostnames of the databases
- Credentials for the databases

To locate database information:

- **Sqoop, Oozie, and Hue** – Go to **Cluster Name > Configuration > Database Settings**.
- **Hive Metastore** – Go to the Hive service, select **Configuration**, and select the **Hive Metastore Database** category.
- **Sentry** – Go to the Sentry service, select **Configuration**, and select the **Sentry Server Database** category.

Step 2: Complete Pre-Upgrade Steps

1. Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) for the new versions you are upgrading to.
2. Read the [CDH 5 Release Notes](#).
3. Read the [Cloudera Security Bulletins](#).
4. Ensure that Java 1.7 or 1.8 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130 or [Upgrading to Oracle JDK 1.8](#) on

page 132, and make sure you have read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.

5. Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

- **Targeted CDH Version:** 5.0.5

Minimum Cloudera Manager Version: 5.0.x

- **Targeted CDH Version:** 5.8.2

Minimum Cloudera Manager Version: 5.8.x

6. If you are upgrading from CDH 5.1 or lower, and use Hive Date partition columns, you may need to update the date format. See [Date partition columns](#).
7. If the cluster uses Impala, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions, or in case of any problems, check against the full list of [Impala keywords](#).
8. Run the [Host Inspector](#) and fix every issue. Go to **Cluster > Inspect Hosts**.
9. Run the [Security Inspector](#) and fix and reported errors. (Go to **Administration > Security** and click **Security Inspector**.)
- 10 Log in to any cluster node as the `hdfs` user, run the following commands, and correct any reported errors:

```
hdfs fsck /
```



Note: the `fsck` command may take 10 minutes or more to complete, depending on the number of files in your cluster.

```
hdfs dfsadmin -report
```

See [HDFS Commands Guide](#) in the Apache Hadoop documentation.

- 11 Log in to any DataNode as the `hbase` user, run the following command, and correct any reported errors:

```
hbase hbck
```

See [Checking and Repairing HBase Tables](#).

- 12 Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- 13 To avoid unnecessary alerts during the upgrade process, enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Be sure to exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.
- 14 If you have configured Hue to use TLS/SSL and you are upgrading from CDH 5.2 or lower to CDH 5.3 or higher, Hue validates CA certificates and requires a truststore. To create a truststore, follow the instructions in [Hue as a TLS/SSL Client](#).
- 15 If your cluster uses the Flume Kafka client, and you are upgrading to CDH 5.8.0 or CDH 5.8.1, perform the extra steps described in [Upgrading to CDH 5.8.0 or CDH 5.8.1 When Using the Flume Kafka Client](#) on page 81 and then continue with the procedures in this topic.
- 16 If your cluster uses Impala and Llama, this role has been deprecated as of CDH 5.9 and you must remove the role from the Impala service before starting the upgrade. If you do not remove this role, the upgrade wizard will halt the upgrade.

To determine if Impala uses Llama:

1. Go to the Impala service.
2. Select the **Instances** tab.

3. Examine the list of roles in the **Role Type** column. If Llama appears, the Impala service is using Llama.

To remove the Llama role:

1. Go to the Impala service and select **Actions > Disable YARN and Impala Integrated Resource Management**.

The **Disable YARN and Impala Integrated Resource Management** wizard displays.

2. Click **Continue**.

The **Disable YARN and Impala Integrated Resource Management Command** page displays the progress of the commands to disable the role.

3. When the commands have completed, click **Finish**.

Step 3: Ensure High Availability Is Enabled

See [HDFS High Availability](#) for instructions. Enabling automatic failover is optional. Automatic failover does not affect the rolling restart operation. If you have JobTracker high availability configured, Cloudera Manager will fail over the JobTracker during the rolling restart, but configuring JobTracker high availability is not a requirement for performing a rolling upgrade.

Step 4: Back Up HDFS Metadata

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

Back up HDFS metadata using the following command:

```
hdfs dfsadmin -fetchImage local directory
```

Step 5: Back Up Databases



Note: Backing up databases requires that you stop some services, which may make them unavailable during backup.



Back up the databases for any of the following services that are deployed in your cluster:

Table 1: Service Databases to Back Up


Service	Where to find database information
Sqoop	Go to Clusters > Cluster Name > Sqoop service > Configuration and select the Database category.
Hue	Go to Clusters > Cluster Name > Hue service > Configuration and select the Database category.
Oozie	Go to Clusters > Cluster Name > Oozie service > Configuration and select the Database category.
Cloudera Navigator Audit Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Cloudera Navigator Metadata Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.

Service	Where to find database information
Activity Monitor	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Reports Manager	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Sentry Server	Go to Clusters > Cluster Name > Sentry service > Configuration and select the Sentry Server Database category.
Hive Metastore	Go to Clusters > Cluster Name > Hive service > Configuration and select the Hive Metastore Database category.

To back up the databases:

- If not already stopped, stop the service:
 - On the **Home > Status** tab, click  to the right of the service name and select **Stop**.
 - Click **Stop** in the next screen to confirm. When you see a **Finished** status, the service has stopped.
- Back up the database. See [Backing Up Databases](#) for detailed instructions for each supported type of database.
- Restart the service:
 - On the **Home > Status** tab, click  to the right of the service name and select **Start**.
 - Click **Start** that appears in the next screen to confirm. When you see a **Finished** status, the service has started.

Step 6: Run the Upgrade Wizard

- From the **Home > Status** tab, click  next to the cluster name and select **Upgrade Cluster**.
The **Getting Started** page of the upgrade wizard displays.
- If the option to pick between packages and parcels displays, select **Use Parcels**.
- In the **Choose CDH Version (Parcels)** field, select the CDH version. If no qualifying parcels are listed, or you want to upgrade to a different version, click the **Modify the Remote Parcel Repository URLs** link to go to the configuration page for **Remote Parcel Repository URLs** and add the appropriate URL to the configuration. See [Parcel Configuration Settings](#) for information about entering the correct URL for parcel repositories. Click **Continue**.
- Read the notices for steps you must complete before upgrading, click the **Yes, I ...** checkboxes after completing the steps, and click **Continue**.
Cloudera Manager verifies that hosts have the correct software installed.
- Click **Continue**.
The selected parcels are downloaded and distributed.
- Click **Continue**.
The Host Inspector runs and displays the CDH version on the hosts.
- Click **Continue**.
The **Choose Upgrade Procedure** screen displays.

8. Select **Rolling Restart**. Cloudera Manager upgrades services and performs a rolling restart. This option is only available if you have [enabled high availability for HDFS](#). Services that do not support rolling restart undergo a normal restart, and are not available during the restart process.
9. (Optional) Configure the following parameters for the rolling restart:

Batch Size

Number of roles to include in a batch. Cloudera Manager restarts the worker roles rack-by-rack, in alphabetical order, and within each rack, hosts are restarted in alphabetical order. If you use the default replication factor of 3, Hadoop tries to keep the replicas on at least 2 different racks. So if you have multiple racks, you can use a higher batch size than the default 1. However, using a batch size that is too high means that fewer worker roles are active at any time during the upgrade, which can cause temporary performance degradation. If you are using a single rack, restart *one worker node at a time* to ensure data availability during upgrade.

Advanced Options > Sleep between batches

Amount of time Cloudera Manager waits before starting the next batch.

Advanced Options > Failed threshold

The number of *batch* failures that cause the entire rolling restart to fail. For example if you have a very large cluster, you can use this option to allow some failures when you know that the cluster is functional when some worker roles are down.

- 10 Click **Continue**.

The **Upgrade Cluster Command** screen displays the result of the commands run by the wizard as it shuts down services, activates the new parcel, upgrades services, deploys client configuration files, restarts services, and performs a rolling restart of the services that support it.

If your cluster was previously installed or upgraded using *packages*, the wizard may indicate that some services cannot start because their parcels are not available. To download the required parcels:

1. In another browser tab, open the Cloudera Manager Admin Console.
2. Select **Hosts > Parcels**.
3. Locate the row containing the missing parcel and click the button to **Download, Distribute**, and then **Activate** the parcel.
4. Return to the upgrade wizard and click the **Retry** button.

The Upgrade Wizard continues upgrading the cluster.

- 11 Click **Finish** to return to the Home page.

Step 7: Recover from Failed Steps or Perform a Manual Upgrade

The actions performed by the upgrade wizard are listed in [Performing Upgrade Wizard Actions Manually](#) on page 80. If any of the steps in the **Upgrade Cluster Command** screen fail, complete the steps as described in that section before proceeding.

Step 8: Remove the Previous CDH Version Packages and Refresh Symlinks

[Not required for CDH maintenance release upgrades.]

Skip this step if your previous installation or upgrade used parcels.

If your previous installation of CDH was done using packages, remove those packages on all hosts where you installed the parcels and refresh the symlinks so that clients will run the new software versions.

1. If your Hue service uses the embedded SQLite database, back up `/var/lib/hue/desktop.db` to a location that is not `/var/lib/hue` because this directory is removed when the packages are removed.
2. Uninstall the CDH packages on each host:
 - **Not including Impala and Search**

RHEL

```
sudo yum remove bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

SLES

```
sudo zypper remove bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

Ubuntu or Debian

```
sudo apt-get purge bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

- **Including Impala and Search**

RHEL

```
sudo yum remove 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

SLES

```
sudo zypper remove 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

Ubuntu or Debian

```
sudo apt-get purge 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

3. Restart all the Cloudera Manager Agents to force an update of the symlinks to point to the newly installed components on each host:

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

4. If your Hue service uses the embedded SQLite database, restore the database you backed up:
 - a. Stop the Hue service.
 - b. Copy the backup from the temporary location to the newly created Hue database directory, `/var/lib/hue`.
 - c. Start the Hue service.

Step 9: Finalize HDFS Rolling Upgrade

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

To determine if you can finalize, run important workloads and ensure that they are successful. Once you have finalized the upgrade, you cannot roll back to a previous version of HDFS without using backups. Verifying that you are ready to finalize the upgrade can take a long time.

1. Go to the HDFS service.
2. Select **Actions > Finalize Rolling Upgrade** and click **Finalize Rolling Upgrade** to confirm.

Step 10: Exit Maintenance Mode

If you entered maintenance mode during this upgrade, [exit maintenance mode](#).

Upgrading to CDH 5.x Using Parcels

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

This topic describes how to upgrade CDH from any version of CDH 5.x to a higher version of CDH 5.x using Cloudera Manager and [parcels](#). If the CDH 5 cluster you are upgrading was installed using packages, you can upgrade it using parcels, and the upgraded version of CDH will then use parcels for future upgrades or changes. You can also [migrate your cluster from using packages to using parcels](#) before starting the upgrade. The minor version of Cloudera Manager you use to perform the upgrade must be equal to or greater than the CDH minor version. To upgrade Cloudera Manager, see [Overview of Upgrading Cloudera Manager](#) on page 9.

The upgrade procedure described in this topic requires cluster downtime. If the cluster was installed using parcels, has a Cloudera Enterprise license, and has HDFS high availability enabled, you can perform a [rolling upgrade](#) that does not require cluster downtime.



Note: If you are upgrading to a *maintenance* version of CDH, skip any steps that are labeled **[Not required for CDH maintenance release upgrades.]**.

The version numbers for maintenance releases differ only in the third digit, for example when upgrading from CDH 5.8.0 to CDH 5.8.2. See [Maintenance Version Upgrades](#) on page 54.

To upgrade CDH using parcels:

Step 1: Collect Upgrade Information

Before starting an upgrade, collect the following information:

1. Host credentials. You must have SSH access and be able to log in using a root account or an account that has password-less sudo permission.
2. The version of Cloudera Manager used in your cluster. Go to **Support > About**.
3. The version of the JDK deployed in the cluster. Go to **Support > About**.
4. The version of CDH. The CDH version number displays next to the cluster name on the **Home** page.
5. Whether the cluster was installed using parcels or packages. This information displays next to the CDH version on the **Home** page of Cloudera Manager.
6. The services enabled in your cluster. Go to **Clusters > Cluster name**.
7. Operating system type and version. Go to **Hosts** and click on a hostname in the list. The operating system type and version displays in the **Distribution** row in the **Details** section.
8. Database information for the databases used by Sqoop, Oozie, Hue, Hive Metastore, and Sentry Server (information is only required if these services are enabled in the cluster).

Gather the following information:

- Type of database (PostgreSQL, Embedded PostgreSQL, MySQL, MariaDB, or Oracle)
- Hostnames of the databases
- Credentials for the databases

To locate database information:

- **Sqoop, Oozie, and Hue** – Go to **Cluster Name > Configuration > Database Settings**.
- **Hive Metastore** – Go to the Hive service, select **Configuration**, and select the **Hive Metastore Database** category.
- **Sentry** – Go to the Sentry service, select **Configuration**, and select the **Sentry Server Database** category.

Step 2: Complete Pre-Upgrade Steps

1. Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) for the new versions you are upgrading to.

2. Read the [CDH 5 Release Notes](#).
3. Read the [Cloudera Security Bulletins](#).
4. Ensure that Java 1.7 or 1.8 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130 or [Upgrading to Oracle JDK 1.8](#) on page 132, and make sure you have read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.
5. Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:
 - **Targeted CDH Version:** 5.0.5
Minimum Cloudera Manager Version: 5.0.x
 - **Targeted CDH Version:** 5.8.2
Minimum Cloudera Manager Version: 5.8.x
6. If you are upgrading from CDH 5.1 or lower, and use Hive Date partition columns, you may need to update the date format. See [Date partition columns](#).
7. If the cluster uses Impala, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions, or in case of any problems, check against the full list of [Impala keywords](#).
8. Run the [Host Inspector](#) and fix every issue. Go to **Cluster > Inspect Hosts**.
9. Run the [Security Inspector](#) and fix and reported errors. (Go to **Administration > Security** and click **Security Inspector**.)
- 10 Log in to any cluster node as the `hdfs` user, run the following commands, and correct any reported errors:

```
hdfs fsck /
```



Note: the `fsck` command may take 10 minutes or more to complete, depending on the number of files in your cluster.

```
hdfs dfsadmin -report
```

See [HDFS Commands Guide](#) in the Apache Hadoop documentation.

- 11 Log in to any DataNode as the `hbase` user, run the following command, and correct any reported errors:

```
hbase hbck
```

See [Checking and Repairing HBase Tables](#).

- 12 Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- 13 To avoid unnecessary alerts during the upgrade process, enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Be sure to exit maintenance mode when you have finished the upgrade to reenale Cloudera Manager alerts.
- 14 If you have configured Hue to use TLS/SSL and you are upgrading from CDH 5.2 or lower to CDH 5.3 or higher, Hue validates CA certificates and requires a truststore. To create a truststore, follow the instructions in [Hue as a TLS/SSL Client](#).
- 15 If your cluster uses the Flume Kafka client, and you are upgrading to CDH 5.8.0 or CDH 5.8.1, perform the extra steps described in [Upgrading to CDH 5.8.0 or CDH 5.8.1 When Using the Flume Kafka Client](#) on page 81 and then continue with the procedures in this topic.
- 16 If your cluster uses Impala and Llama, this role has been deprecated as of CDH 5.9 and you must remove the role from the Impala service before starting the upgrade. If you do not remove this role, the upgrade wizard will halt the upgrade.

To determine if Impala uses Llama:

1. Go to the Impala service.
2. Select the **Instances** tab.
3. Examine the list of roles in the **Role Type** column. If Llama appears, the Impala service is using Llama.

To remove the Llama role:

1. Go to the Impala service and select **Actions > Disable YARN and Impala Integrated Resource Management**.

The **Disable YARN and Impala Integrated Resource Management** wizard displays.

2. Click **Continue**.

The **Disable YARN and Impala Integrated Resource Management Command** page displays the progress of the commands to disable the role.

3. When the commands have completed, click **Finish**.

Step 3: Stop Cluster Services

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Stop**.

2. Click **Stop** in the confirmation screen. The **Command Details** window shows the progress of stopping services.

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

Step 4: Back up the HDFS Metadata on the NameNode

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

1. Go to the HDFS service.
2. Click the **Configuration** tab.
3. In the Search field, search for "NameNode Data Directories" and note the value.
4. On the active NameNode host, back up the directory listed in the NameNode Data Directories property. If more than one is listed, make a backup of one directory, because each directory is a complete copy. For example, if the NameNode data directory is `/data/dfs/nn`, do the following as root:

```
# cd /data/dfs/nn
# tar -cvf /root/nn_backup_data.tar .
```

You should see output like this:

```
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
```



```
./current/edits
./image/
./image/fsimage
```

If a file with the extension *lock* exists in the NameNode data directory, the NameNode most likely is still running. Repeat the steps, beginning with shutting down the NameNode role.

Step 5: Back Up Databases





Note: Backing up databases requires that you stop some services, which may make them unavailable during backup.

Back up the databases for any of the following services that are deployed in your cluster:

Table 2: Service Databases to Back Up


Service	Where to find database information
Sqoop	Go to Clusters > <i>Cluster Name</i> > <i>Sqoop service</i> > Configuration and select the Database category.
Hue	Go to Clusters > <i>Cluster Name</i> > <i>Hue service</i> > Configuration and select the Database category.
Oozie	Go to Clusters > <i>Cluster Name</i> > <i>Oozie service</i> > Configuration and select the Database category.
Cloudera Navigator Audit Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Cloudera Navigator Metadata Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Activity Monitor	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Reports Manager	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Sentry Server	Go to Clusters > <i>Cluster Name</i> > <i>Sentry service</i> > Configuration and select the Sentry Server Database category.
Hive Metastore	Go to Clusters > <i>Cluster Name</i> > <i>Hive service</i> > Configuration and select the Hive Metastore Database category.

To back up the databases:

- If not already stopped, stop the service:
 - On the **Home** > **Status** tab, click  to the right of the service name and select **Stop**.
 - Click **Stop** in the next screen to confirm. When you see a **Finished** status, the service has stopped.
- Back up the database. See [Backing Up Databases](#) for detailed instructions for each supported type of database.
- Restart the service:
 - On the **Home** > **Status** tab, click  to the right of the service name and select **Start**.

- b. Click **Start** that appears in the next screen to confirm. When you see a **Finished** status, the service has started.

Step 6: Run the Upgrade Wizard

1. From the **Home > Status** tab, click  next to the cluster name and select **Upgrade Cluster**.
The **Getting Started** page of the upgrade wizard displays.
2. If the option to pick between packages and parcels displays, select **Use Parcels**.
3. In the **Choose CDH Version (Parcels)** field, select the CDH version. If no qualifying parcels are listed, or you want to upgrade to a different version, click the **Modify the Remote Parcel Repository URLs** link to go to the configuration page for **Remote Parcel Repository URLs** and add the appropriate URL to the configuration. See [Parcel Configuration Settings](#) for information about entering the correct URL for parcel repositories. Click **Continue**.
4. Read the notices for steps you must complete before upgrading, click the **Yes, I ...** checkboxes after completing the steps, and click **Continue**.

Cloudera Manager verifies that hosts have the correct software installed.

5. Click **Continue**.

The Host Inspector runs and displays the CDH version on the hosts.

6. Click **Continue**.

The **Choose Upgrade Procedure** screen displays the available types of upgrades:

- **Full Cluster Restart** - Cloudera Manager performs all service upgrades and restarts the cluster.
- **Manual upgrade** Cloudera Manager configures the cluster to the specified CDH version but performs no upgrades or service restarts. Manually upgrading is difficult and for advanced users only. To perform a manual upgrade:
 1. Select the **Let me upgrade the cluster** checkbox.
 2. Click **Continue**.
 3. See [Performing Upgrade Wizard Actions Manually](#) on page 80 for the required steps.

7. Select **Full Cluster Restart**.

8. Click **Continue**.

The **Upgrade Cluster Command** screen displays the result of the commands run by the wizard as it shuts down all services, activates the new parcel, upgrades services, deploys client configuration files, and restarts services. If any of the steps fail, correct any reported errors and click the **Retry** button. If you click the **Abort** button, the **Retry** button at the top right is enabled.

Click **Retry** to retry the step and continue the wizard, or click the Cloudera Manager logo to return to the **Home > Status** tab and [manually perform the failed step and all following steps](#).

9. Click **Continue**.

The wizard reports the result of the upgrade.

If your cluster was previously installed or upgraded using *packages*, the wizard may indicate that some services cannot start because their parcels are not available. To download the required parcels:

1. In another browser tab, open the Cloudera Manager Admin Console.
2. Select **Hosts > Parcels**.
3. Locate the row containing the missing parcel and click the button to **Download, Distribute**, and then **Activate** the parcel.
4. Return to the upgrade wizard and click the **Retry** button.

The Upgrade Wizard continues upgrading the cluster.

10 Click **Finish** to return to the Home page.

Step 7: Recover from Failed Steps or Perform a Manual Upgrade

The actions performed by the upgrade wizard are listed in [Performing Upgrade Wizard Actions Manually](#) on page 80. If any of the steps in the **Upgrade Cluster Command** screen fail, complete the steps as described in that section before proceeding.

Step 8: Remove the Previous CDH Version Packages and Refresh Symlinks

[Not required for CDH maintenance release upgrades.]

Skip this step if your previous installation or upgrade used parcels.

If your previous installation of CDH was done using packages, remove those packages on all hosts where you installed the parcels and refresh the symlinks so that clients will run the new software versions.

1. If your Hue service uses the embedded SQLite database, back up `/var/lib/hue/desktop.db` to a location that is not `/var/lib/hue` because this directory is removed when the packages are removed.
2. Uninstall the CDH packages on each host:

- **Not including Impala and Search**

RHEL

```
sudo yum remove bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

SLES

```
sudo zypper remove bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

Ubuntu or Debian

```
sudo apt-get purge bigtop-utils bigtop-jsvc bigtop-tomcat hue-common sqoop2-client
```

- **Including Impala and Search**

RHEL

```
sudo yum remove 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

SLES

```
sudo zypper remove 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

Ubuntu or Debian

```
sudo apt-get purge 'bigtop-*' hue-common impala-shell solr-server sqoop2-client
hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc
```

3. Restart all the Cloudera Manager Agents to force an update of the symlinks to point to the newly installed components on each host:

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

4. If your Hue service uses the embedded SQLite database, restore the database you backed up:
 - a. Stop the Hue service.
 - b. Copy the backup from the temporary location to the newly created Hue database directory, `/var/lib/hue`.
 - c. Start the Hue service.

Step 9: Finalize the HDFS Metadata Upgrade

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

To determine if you can finalize, run important workloads and ensure that they are successful. Once you have finalized the upgrade, you cannot roll back to a previous version of HDFS without using backups. Verifying that you are ready to finalize the upgrade can take a long time.

Make sure you have enough free disk space, keeping in mind that the following behavior continues until the upgrade is finalized:

- Deleting files does not free up disk space.
- Using the balancer causes all moved replicas to be duplicated.
- All on-disk data representing the NameNodes metadata is retained, which could more than double the amount of space required on the NameNode and JournalNode disks.

To finalize the metadata upgrade:

1. Go to the HDFS service.
2. Click the **Instances** tab.
3. Select the **NameNode** instance. If you have enabled high availability for HDFS, select **NameNode (Active)**.
4. Select **Actions > Finalize Metadata Upgrade** and click **Finalize Metadata Upgrade** to confirm.

Step 10: Exit Maintenance Mode

If you entered maintenance mode during this upgrade, [exit maintenance mode](#).

Upgrading to CDH 5.x Using Packages

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

This topic describes how to upgrade CDH from any version of CDH 5.x to a higher version of CDH 5.x using Cloudera Manager and *packages*. The minor version of Cloudera Manager you use to perform the upgrade must be equal to or greater than the CDH minor version. To upgrade Cloudera Manager, see [Overview of Upgrading Cloudera Manager](#) on page 9.

The upgrade procedure described in this topic requires cluster downtime. If the cluster was installed using parcels, has a Cloudera Enterprise license, and has HDFS high availability enabled, you can perform a [rolling upgrade](#) that does not require cluster downtime.

If the CDH 5 cluster you are upgrading was installed using packages, you can [upgrade it using parcels](#), and the upgraded version of CDH will then use parcels for future upgrades or changes. You can also [migrate your cluster from using packages to using parcels](#) before starting the upgrade.

Upgrading using packages is for advanced users. Some parts of this procedure require you to run commands from the command line on *all cluster hosts*. This may require significant time. Additionally, if other software packages are installed on cluster hosts, the upgrade steps described in this topic could update dependencies that affect those packages.



Note: If you are upgrading to a *maintenance* version of CDH, skip any steps that are labeled **[Not required for CDH maintenance release upgrades.]**.

The version numbers for maintenance releases differ only in the third digit, for example when upgrading from CDH 5.8.0 to CDH 5.8.2. See [Maintenance Version Upgrades](#) on page 54.

To upgrade CDH using packages:

Step 1: Collect Upgrade Information

Before starting an upgrade, collect the following information:

1. Host credentials. You must have SSH access and be able to log in using a root account or an account that has password-less sudo permission.
2. The version of Cloudera Manager used in your cluster. Go to **Support > About**.
3. The version of the JDK deployed in the cluster. Go to **Support > About**.
4. The version of CDH. The CDH version number displays next to the cluster name on the **Home** page.
5. Whether the cluster was installed using parcels or packages. This information displays next to the CDH version on the **Home** page of Cloudera Manager.
6. The services enabled in your cluster. Go to **Clusters > Cluster name**.
7. Operating system type and version. Go to **Hosts** and click on a hostname in the list. The operating system type and version displays in the **Distribution** row in the **Details** section.
8. Database information for the databases used by Sqoop, Oozie, Hue, Hive Metastore, and Sentry Server (information is only required if these services are enabled in the cluster).

Gather the following information:

- Type of database (PostgreSQL, Embedded PostgreSQL, MySQL, MariaDB, or Oracle)
- Hostnames of the databases
- Credentials for the databases

To locate database information:

- **Sqoop, Oozie, and Hue** – Go to **Cluster Name > Configuration > Database Settings**.
- **Hive Metastore** – Go to the Hive service, select **Configuration**, and select the **Hive Metastore Database** category.
- **Sentry** – Go to the Sentry service, select **Configuration**, and select the **Sentry Server Database** category.

Step 2: Complete Pre-Upgrade Steps

1. Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](#) for the new versions you are upgrading to.
2. Read the [CDH 5 Release Notes](#).
3. Read the [Cloudera Security Bulletins](#).
4. Ensure that Java 1.7 or 1.8 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130 or [Upgrading to Oracle JDK 1.8](#) on page 132, and make sure you have read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.
5. Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:
 - **Targeted CDH Version:** 5.0.5
Minimum Cloudera Manager Version: 5.0.x
 - **Targeted CDH Version:** 5.8.2
Minimum Cloudera Manager Version: 5.8.x
6. If you are upgrading from CDH 5.1 or lower, and use Hive Date partition columns, you may need to update the date format. See [Date partition columns](#).
7. If the cluster uses Impala, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions, or in case of any problems, check against the full list of [Impala keywords](#).
8. Run the [Host Inspector](#) and fix every issue. Go to **Cluster > Inspect Hosts**.
9. Run the [Security Inspector](#) and fix and reported errors. (Go to **Administration > Security** and click **Security Inspector**.)

- 10 Log in to any cluster node as the `hdfs` user, run the following commands, and correct any reported errors:

```
hdfs fsck /
```



Note: the `fsck` command may take 10 minutes or more to complete, depending on the number of files in your cluster.

```
hdfs dfsadmin -report
```

See [HDFS Commands Guide](#) in the Apache Hadoop documentation.

- 11 Log in to any DataNode as the `hbase` user, run the following command, and correct any reported errors:

```
hbase hbck
```

See [Checking and Repairing HBase Tables](#).

- 12 Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- 13 To avoid unnecessary alerts during the upgrade process, enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Be sure to exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.
- 14 If you have configured Hue to use TLS/SSL and you are upgrading from CDH 5.2 or lower to CDH 5.3 or higher, Hue validates CA certificates and requires a truststore. To create a truststore, follow the instructions in [Hue as a TLS/SSL Client](#).
- 15 If your cluster uses the Flume Kafka client, and you are upgrading to CDH 5.8.0 or CDH 5.8.1, perform the extra steps described in [Upgrading to CDH 5.8.0 or CDH 5.8.1 When Using the Flume Kafka Client](#) on page 81 and then continue with the procedures in this topic.
- 16 If your cluster uses Impala and Llama, this role has been deprecated as of CDH 5.9 and you must remove the role from the Impala service before starting the upgrade. If you do not remove this role, the upgrade wizard will halt the upgrade.

To determine if Impala uses Llama:

1. Go to the Impala service.
2. Select the **Instances** tab.
3. Examine the list of roles in the **Role Type** column. If Llama appears, the Impala service is using Llama.

To remove the Llama role:

1. Go to the Impala service and select **Actions > Disable YARN and Impala Integrated Resource Management**. The **Disable YARN and Impala Integrated Resource Management** wizard displays.
2. Click **Continue**.
The **Disable YARN and Impala Integrated Resource Management Command** page displays the progress of the commands to disable the role.
3. When the commands have completed, click **Finish**.

Step 3: Upgrade Unmanaged Components

Upgrade any unmanaged components before upgrading components that are managed by Cloudera Manager.

Step 4: Stop Cluster Services

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Stop**.

2. Click **Stop** in the confirmation screen. The **Command Details** window shows the progress of stopping services.

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

Step 5: Back up the HDFS Metadata on the NameNode

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

1. Go to the HDFS service.
2. Click the **Configuration** tab.
3. In the Search field, search for "NameNode Data Directories" and note the value.
4. On the active NameNode host, back up the directory listed in the NameNode Data Directories property. If more than one is listed, make a backup of one directory, because each directory is a complete copy. For example, if the NameNode data directory is `/data/dfs/nn`, do the following as root:

```
# cd /data/dfs/nn
# tar -cvf /root/nn_backup_data.tar .
```

You should see output like this:

```
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
./image/
./image/fsimage
```

If a file with the extension *lock* exists in the NameNode data directory, the NameNode most likely is still running. Repeat the steps, beginning with shutting down the NameNode role.

Step 6: Back Up Databases



Note: Backing up databases requires that you stop some services, which may make them unavailable during backup.

Back up the databases for any of the following services that are deployed in your cluster:

Table 3: Service Databases to Back Up

Service	Where to find database information
Sqoop	Go to Clusters > Cluster Name > Sqoop service > Configuration and select the Database category.
Hue	Go to Clusters > Cluster Name > Hue service > Configuration and select the Database category.

Service	Where to find database information
Oozie	Go to Clusters > Cluster Name > Oozie service > Configuration and select the Database category.
Cloudera Navigator Audit Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Cloudera Navigator Metadata Server	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Activity Monitor	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Reports Manager	Go to Clusters > Cloudera Management Service > Configuration and select the Database category.
Sentry Server	Go to Clusters > Cluster Name > Sentry service > Configuration and select the Sentry Server Database category.
Hive Metastore	Go to Clusters > Cluster Name > Hive service > Configuration and select the Hive Metastore Database category.

To back up the databases:

1. If not already stopped, stop the service:

- a. On the **Home** > **Status** tab, click



to the right of the service name and select **Stop**.

- b. Click **Stop** in the next screen to confirm. When you see a **Finished** status, the service has stopped.

2. Back up the database. See [Backing Up Databases](#) for detailed instructions for each supported type of database.

3. Restart the service:

- a. On the **Home** > **Status** tab, click



to the right of the service name and select **Start**.

- b. Click **Start** that appears in the next screen to confirm. When you see a **Finished** status, the service has started.

Step 7: Upgrade Managed Components

Use *one* of the following strategies to upgrade CDH 5:

Use the Cloudera 1-Click Package

Using the Cloudera "1-click Install" package is the simplest way to upgrade only the Cloudera packages.

1. Check whether you have the CDH 5 "1-click" repository installed by running the following command on each cluster host:

RHEL/CentOS-compatible and SLES

```
rpm -q CDH 5-repository
```

If you are upgrading from CDH 5 Beta 1 or higher, and you used the "1-click" package for the previous CDH 5 release, you should see:

```
CDH5-repository-1-0
```


In this case, skip to Step 3 on page 74, **Install the CDH packages**. If instead you see:

```
package CDH 5-repository is not installed
```

proceed with Step 2 on page 73, **Install the 1-click package**.

Ubuntu and Debian

```
dpkg -l | grep CDH 5-repository
```

If the repository is installed, skip to Step 3 on page 74, **Install the CDH packages**; otherwise proceed with Step 2 on page 73, **Install the 1-click package**.

2. **Install the CDH5 "1-click" package.** If the CDH 5 "1-click" repository is not already installed on each host in the cluster, follow the instructions below for that host's operating system.

RHEL compatible

1. Download and install the "1-click Install" package on each cluster host.

- a. Download the CDH 5 "1-click Install" package (or RPM).

Click the appropriate RPM and **Save File** to a directory with write access (for example, your home directory).

OS Version	Link to CDH 5 RPM
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS/Oracle 6	RHEL/CentOS/Oracle 6 link
RHEL/CentOS/Oracle 7	RHEL/CentOS/Oracle 7 link

- b. Install the RPM for all RHEL versions:

```
$ sudo yum --nogpgcheck localinstall cloudera-cdh-5-0.x86_64.rpm
```

2. Optionally add a repository key. Using a repository key helps to verify that you are using a signed version of the package. If you add a repository key, you can omit the `--nogpgcheck` option when running `yum` commands. Run the following command to add the repository key:

Red Hat/CentOS/Oracle 5

```
$ sudo rpm --import
http://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

Red Hat/CentOS/Oracle 6

```
$ sudo rpm --import
http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

SLES

1. Download and install the "1-click Install" package on each cluster host:

- a. Download the CDH 5 "1-click Install" package.

Download the [RPM file](#), choose **Save File**, and save it to a directory to which you have write access (for example, your home directory).

- b. Install the RPM:

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

- c. Update your system package index by running the following:

```
$ sudo zypper refresh
```

2. Optionally add a repository key. Using a repository key helps to verify that you are using a signed version of the package.

```
$ sudo rpm --import  
http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

Ubuntu and Debian

1. Download and install the "1-click Install" package on each cluster host:

- a. Download the CDH 5 "1-click Install" package:

OS Version	Package Link
Jessie	Jessie package
Wheezy	Wheezy package
Precise	Precise package
Trusty	Trusty package

- b. Install the package by doing one of the following:

- Choose **Open with** in the download window to use the package manager.
- Choose **Save File**, save the package to a directory to which you have write access (for example, your home directory), and install it from the command line. For example:

```
sudo dpkg -i cdh5-repository_1.0_all.deb
```

2. (Optionally) add a repository key (Using a repository key helps to verify that you are using a signed version of the package.):

- **Ubuntu Trusty**

```
$ curl -s http://archive.cloudera.com/cdh5/ubuntu/trusty/amd64/cdh/archive.key | sudo  
apt-key add -
```

- **Ubuntu Precise**

```
$ curl -s http://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key | sudo  
apt-key add -
```

- **Debian Wheezy**

```
$ curl -s http://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key | sudo  
apt-key add -
```

3. Install the CDH packages by running the following command on all cluster hosts:

RHEL compatible

```
$ sudo yum clean all  
$ sudo yum install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-httpfs hadoop-kms  
hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala hue-pig  
hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala impala-shell
```

```
kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr solr-mapreduce
spark-python sqoop sqoop2 whirr zookeeper
```

SLES

```
$ sudo zypper clean --all
$ sudo zypper install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-httpfs hadoop-kms
hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala hue-pig
hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala impala-shell
kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr solr-mapreduce
spark-python sqoop sqoop2 whirr zookeeper
```

Ubuntu and Debian

```
$ sudo apt-get update
$ sudo apt-get install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-httpfs
hadoop-kms hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala
hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala
impala-shell kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr
solr-mapreduce spark-python sqoop sqoop2 whirr zookeeper
```



Note: Installing these packages also installs all the other CDH packages required for a full CDH 5 installation.

Use Operating System Package Management Tools

- Use your operating system's package management tools to update all packages to the latest version using standard repositories. This approach minimizes the amount of configuration required and uses the simplest commands. This can take a considerable amount of time if you have not upgraded the system recently. To update all packages on your system, run the following command on each cluster host:

RHEL

```
$ sudo yum update
```

SLES

```
$ sudo zypper up
```

Ubuntu or Debian

```
$ sudo apt-get upgrade
```

Use a Specific Set of Packages

To upgrade managed components to a specific version of CDH, specify the packages you want to use for the upgrade. Follow the procedure at [Upgrade Managed Components Using a Specific Set of Packages](#) on page 77 and then continue with the procedures in this topic.


Step 8: Update Symlinks for the Newly Installed Components

Restart the Cloudera Manager Agents to force an update of the symlinks to point to the newly installed components on each host. Run the following command on all cluster hosts:

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

Step 9: Run the Upgrade Wizard

1. Log in to the Cloudera Manager Admin console.

2. From the **Home > Status** tab, click  next to the cluster name and select **Upgrade Cluster**. The Upgrade Wizard starts.
3. In the **Choose Method** field, select the **Use Packages** option.
4. In the **Choose CDH Version (Packages)** field, specify the CDH version of the packages you have installed on your cluster. Click **Continue**.
5. Read the notices for steps you must complete before upgrading, click the **Yes, I ...** checkboxes after completing the steps, and click **Continue**.
6. Cloudera Manager checks that cluster hosts have the correct software installed. If the packages have not been installed, a warning displays to that effect. Install the missing packages and click **Retry**. When there are no errors, click **Continue**.
7. The Host Inspector runs. Correct any errors displayed and click **Continue**.

The **Choose Upgrade Procedure** screen displays the available types of upgrades:

- **Full Cluster Restart** - Cloudera Manager performs all service upgrades and restarts the cluster.
- **Manual upgrade** Cloudera Manager configures the cluster to the specified CDH version but performs no upgrades or service restarts. Manually upgrading is difficult and for advanced users only. To perform a manual upgrade:
 1. Select the **Let me upgrade the cluster** checkbox.
 2. Click **Continue**.
 3. See [Performing Upgrade Wizard Actions Manually](#) on page 80 for the required steps.

8. Select **Full Cluster Restart**.
9. Click **Continue**. The **Upgrade Cluster Command** screen displays the result of the commands run by the wizard as it shuts down all services, upgrades services, deploys client configuration files, and restarts services. If any of the steps fails or if you click the **Abort** button, the **Retry** button at the top right is enabled.

Click **Retry** to retry the step and continue the wizard, or click the Cloudera Manager logo to return to the **Home > Status** tab and manually perform the failed step and all following steps. See [Performing Upgrade Wizard Actions Manually](#) on page 80.

- 10 Click **Continue**. The wizard reports the result of the upgrade.
- 11 Click **Finish** to return to the Home page.

Step 10: Recover from Failed Steps or Perform a Manual Upgrade

The actions performed by the upgrade wizard are listed in [Performing Upgrade Wizard Actions Manually](#) on page 80. If any of the steps in the **Upgrade Cluster Command** screen fail, complete the steps as described in that section before proceeding.

Step 11: Finalize the HDFS Metadata Upgrade

[Not required for CDH maintenance release upgrades.]

The steps in this section are only required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

To determine if you can finalize, run important workloads and ensure that they are successful. Once you have finalized the upgrade, you cannot roll back to a previous version of HDFS without using backups. Verifying that you are ready to finalize the upgrade can take a long time.

Make sure you have enough free disk space, keeping in mind that the following behavior continues until the upgrade is finalized:

- Deleting files does not free up disk space.
- Using the balancer causes all moved replicas to be duplicated.

- All on-disk data representing the NameNodes metadata is retained, which could more than double the amount of space required on the NameNode and JournalNode disks.

To finalize the metadata upgrade:

1. Go to the HDFS service.
2. Click the **Instances** tab.
3. Select the **NameNode** instance. If you have enabled high availability for HDFS, select **NameNode (Active)**.
4. Select **Actions > Finalize Metadata Upgrade** and click **Finalize Metadata Upgrade** to confirm.

Step 12: Exit Maintenance Mode

If you entered maintenance mode during this upgrade, [exit maintenance mode](#).

Upgrade Managed Components Using a Specific Set of Packages



Important: The procedures in this topic are part of the procedure [Upgrading to CDH 5.x Using Packages](#) on page 68. They describe how to configure Cloudera Manager to upgrade managed CDH components using a specific set of packages instead of using the latest packages available. After completing the steps in this topic, return to [Step 7: Upgrade Managed Components](#) on page 72 to complete the upgrade.

1. Download and save the repo file.

- **RHEL-compatible systems**

Click the entry in the table below that matches your RHEL or CentOS system, go to the repo file for your system, and save it in the `/etc/yum.repos.d/` directory.

OS Version	Link
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS 6 (64-bit)	RHEL/CentOS 6 link

- **SLES systems**

1. Run the following command:

```
$ sudo zypper addrepo -f
http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/cloudera-cdh5.repo
```

2. Update your system package index by running:

```
$ sudo zypper refresh
```

- **Ubuntu and Debian systems**

Create a new file `/etc/apt/sources.list.d/cloudera.list` with the following contents:

- For Ubuntu systems:

```
deb [arch=amd64] http://archive.cloudera.com/cdh5/ <OS-release-arch> <RELEASE>-cdh5
contrib deb-src http://archive.cloudera.com/cdh5/ <OS-release-arch> <RELEASE>-cdh5
contrib
```

- For Debian systems:

```
deb http://archive.cloudera.com/cdh5/ <OS-release-arch> <RELEASE>-cdh5 contrib deb-src
http://archive.cloudera.com/cdh5/ <OS-release-arch> <RELEASE>-cdh5 contrib
```

where: <OS-release-arch> is `debian/wheezy/amd64/cdh` or `ubuntu/precise/amd64/cdh`, and <RELEASE> is the name of your distribution, which you can find by running `lsb_release -c`.

2. Edit the repo file to point to the release you want to install or upgrade to.

- **RHEL-compatible systems**

Open the repo file you just saved and change the 5 at the end of the line that begins `baseurl=` to the version number you want.

For example, if you have saved the file for [RHEL 6](#), it looks like this when you open it:

```
[cloudera-cdh5]
name=Cloudera's Distribution for Hadoop, Version 5
baseurl=http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5/
gpgkey = http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

For example, if you want to install CDH 5.1.0, change

```
baseurl=http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5/
```

to

```
baseurl=http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5.1.0/
```

In this example, the resulting file looks like this:

```
[cloudera-cdh5]
name=Cloudera's Distribution for Hadoop, Version 5
baseurl=http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5.1.0/
gpgkey = http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

- **SLES systems**

Open the repo file that you just added and change the 5 at the end of the line that begins `baseurl=` to the version number you want.

The file should look like this when you open it:

```
[cloudera-cdh5]
name=Cloudera's Distribution for Hadoop, Version 5
baseurl=http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/5/
gpgkey = http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

For example, if you want to install CDH 5.1.0, change

```
baseurl=http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/5/
```

to

```
baseurl= http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/5.1.0/
```

In this example, the resulting file looks like this:

```
[cloudera-cdh5]
name=Cloudera's Distribution for Hadoop, Version 5
baseurl=http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/5.1.0/
gpgkey = http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
gpgcheck = 1
```

- **Ubuntu and Debian systems**

Replace `-cdh5` near the end of each line (before `contrib`) with the CDH release you need to install. The following examples use CDH 5.1.0:

64-bit Ubuntu Precise

```
deb [arch=amd64] http://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh
precise-cdh5.1.0 contrib
deb-src http://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh precise-cdh5.1.0
contrib
```

Debian Wheezy

```
deb http://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh wheezy-cdh5.1.0 contrib
deb-src http://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh wheezy-cdh5.1.0 contrib
```

3. Optionally add a repository key:

- **RHEL-compatible**

- **Red Hat/CentOS/Oracle 5**

```
$ sudo rpm --import
http://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **Red Hat/CentOS/Oracle 6**

```
$ sudo rpm --import
http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **SLES**

```
$ sudo rpm --import
http://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **Ubuntu and Debian**

- **Ubuntu Precise**

```
$ curl -s http://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key | sudo
apt-key add -
```

- **Debian Wheezy**

```
$ curl -s http://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key | sudo
apt-key add -
```

4. Install the CDH packages:

- **RHEL-compatible**

```
$ sudo yum clean all
$ sudo yum install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-httpfs hadoop-kms
hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala hue-pig
hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala impala-shell
kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr solr-mapreduce
spark-python sqoop sqoop2 whirr zookeeper
```

- **SLES**

```
$ sudo zypper clean --all
$ sudo zypper install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-httpfs hadoop-kms
hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala hue-pig
hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala impala-shell
kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr solr-mapreduce
spark-python sqoop sqoop2 whirr zookeeper
```

- **Ubuntu and Debian**

```
$ sudo apt-get update
$ sudo apt-get install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-https
hadoop-kms hbase hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala
hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala
impala-shell kite llama mahout oozie parquet pig pig-udf-datafu search sentry solr
solr-mapreduce spark-python sqoop sqoop2 whirr zookeeper
```



Note: Installing these packages also installs all the other CDH packages required for a full CDH 5 installation.

5. Return to [Step 7: Upgrade Managed Components](#) on page 72 to complete the upgrade.

Performing Upgrade Wizard Actions Manually



Important: Perform the steps in this section only if the upgrade wizard reports a failure.

Upgrade HDFS Metadata

If upgrading from:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.2 or 5.3 to 5.4 or higher

1. Start the ZooKeeper service.
2. Go to the HDFS service.
3. Select **Actions > Upgrade HDFS Metadata** and click **Upgrade HDFS Metadata** to confirm.

Upgrade the Hive Metastore Database

Required for the following upgrades:

- CDH 5.0 or 5.1 to 5.2 or higher
- CDH 5.3 to 5.4 or higher

1. Go to the Hive service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
4. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade the Oozie ShareLib

1. Go to the Oozie service.
2. Select **Actions > Start** and click **Start** to confirm.
3. Select **Actions > Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Upgrade the Sentry Database

Required for the following upgrades:

- CDH 5.1 to 5.2 or higher
 - CDH 5.2 to 5.3 or higher
 - CDH 5.4 to 5.5 or higher
1. Go to the Sentry service.
 2. Select **Actions** > **Stop** and click **Stop** to confirm.
 3. Select **Actions** > **Upgrade Sentry Database Tables** and click **Upgrade Sentry Database Tables** to confirm.

Upgrade Spark

1. Go to the Spark service.
2. Select **Actions** > **Stop** and click **Stop** to confirm.
3. Select **Actions** > **Install Spark JAR** and click **Install Spark JAR** to confirm.

Start Cluster Services

1. On the **Home** > **Status** tab, click




to the right of the cluster name and select **Start**.

2. Click **Start** that appears in the next screen to confirm. The **Command Details** window shows the progress of starting services.

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

Deploy Client Configuration Files

1. On the Home page, click  to the right of the cluster name and select **Deploy Client Configuration**.
2. Click the **Deploy Client Configuration** button in the confirmation pop-up that appears.

Upgrading to CDH 5.8.0 or CDH 5.8.1 When Using the Flume Kafka Client

Due to the change of offset storage from ZooKeeper to Kafka in the CDH 5.8 Flume Kafka client, data might not be consumed by the Flume agents, or might be duplicated (if `kafka.auto.offset.reset=smallest`) during an upgrade to CDH 5.8.0 or CDH 5.8.1. To prevent this, perform the steps described below before you upgrade your system.



Important: This issue has been fixed for CDH 5.8.2 and higher. If you are upgrading to CDH 5.8.2 or higher, you do not need to perform this procedure. For more information, see [Cloudera Distribution of Apache Kafka Known Issues](#).

The upgrade process is based on this example configuration:

```
tier1.sources = source1
tier1.channels = channel1
tier1.sinks = sink1

tier1.sources.source1.type = org.apache.flume.source.kafka.KafkaSource
tier1.sources.source1.zookeeperConnect = zkhost:2181
tier1.sources.source1.topic = flumetopic
tier1.sources.source1.groupId = flume
tier1.sources.source1.channels = channel1
tier1.sources.source1.interceptors = i1 i2
tier1.sources.source1.interceptors.i1.type = timestamp
```

```
tier1.sources.source1.interceptors.i2.type = host
tier1.sources.source1.kafka.consumer.timeout.ms = 100

tier1.channels.channel1.type = org.apache.flume.channel.kafka.KafkaChannel
tier1.channels.channel1.brokerList=broker1:9092,broker2:9092
tier1.channels.channel1.zookeeperConnect=zkhost:2181
tier1.channels.channel1.topic=flumechannel1
tier1.channels.channel1.groupId = flumechannel
tier1.channels.channel1.capacity = 10000
tier1.channels.channel1.transactionCapacity = 1000

tier1.sinks.sink1.type = hdfs
tier1.sinks.sink1.hdfs.path = /tmp/kafka/%{topic}/
tier1.sinks.sink1.hdfs.filePrefix = %{host}-
tier1.sinks.sink1.hdfs.rollInterval = 60
tier1.sinks.sink1.hdfs.rollSize = 0
tier1.sinks.sink1.hdfs.rollCount = 0
tier1.sinks.sink1.hdfs.fileType = DataStream
tier1.sinks.sink1.channel = channel1
tier1.sinks.sink1.hdfs.kerberosKeytab = $KERBEROS_KEYTAB
tier1.sinks.sink1.hdfs.kerberosPrincipal = $KERBEROS_PRINCIPAL
```

Perform the following steps to upgrade CDH.

1. If you are using a version lower than CDH 5.7, first [upgrade to CDH 5.7](#). If for some reason you cannot upgrade to CDH 5.7, contact your Cloudera Sales Engineer for assistance, or file a support case with specific versions from which and to which you are upgrading.
2. Add the following sections to the source and channel for the Flume configuration:

```
# Added for source upgrade compatability
tier1.sources.source1.kafka.bootstrap.servers = broker1:9092,broker2:9092
tier1.sources.source1.kafka.offsets.storage = kafka
tier1.sources.source1.kafka.dual.commit.enabled = true
tier1.sources.source1.kafka.consumer.group.id = flume
tier1.sources.source1.kafka.topics = flumetopic

# Added for channel upgrade compatability
tier1.channels.channel1.kafka.topic = flumechannel1
tier1.channels.channel1.kafka.bootstrap.servers = broker1:9092,broker2:9092
tier1.channels.channel1.kafka.consumer.group.id = flumechannel
tier1.channels.channel1.kafka.offsets.storage = kafka
tier1.channels.channel1.kafka.dual.commit.enabled = true
```

3. Restart (or rolling restart) the Flume agents. This switches `offsets.storage` to Kafka, but keeps both the Kafka and ZooKeeper offsets updated because the `dual.commit.enabled` property is set to `true`. Confirm that Kafka messages are flowing through the Flume servers. Updating the offsets only occurs when new messages are consumed, so there must be at least one Kafka message consumed by the Flume agent, or one event passed through the Flume channel. Use the following commands to verify that Flume is properly updating the offsets in Kafka (the `egrep` command is used to match the correct topic names: in this example, `flumetopic` and `flumechannel1`):

```
echo "exclude.internal.topics=false" > /tmp/consumer.config
kafka-console-consumer --consumer.config /tmp/consumer.config
--formatter "kafka.coordinator.GroupMetadataManager\$OffsetsMessageFormatter"
--zookeeper zkhost:2181 --topic __consumer_offsets |egrep -e "flumetopic|flumechannel1"
```

Output should be similar to the following and show that the Flume source and/or channel topics offsets are being incremented:

```
[flume,flumetopic,0]::[OffsetMetadata[70,cf9e5630-214e-4689-9869-5e077c936ffb],CommitTime
1469827951129,ExpirationTime 1469914351129]
[flumechannel,flumechannel1,0]::[OffsetMetadata[61,875e7a82-1c22-43be-acaa-eb4d63e7f71e],CommitTime
1469827951128,ExpirationTime 1469914351128]
[flumechannel,flumechannel1,0]::[OffsetMetadata[62,66bda888-0a70-4a02-a286-7e2e7d14050d],CommitTime
1469827951131,ExpirationTime 1469914351131]
```

4. Perform the upgrade to CDH 5.8.0 or CDH 5.8.1. The Flume agents are restarted during the process. Flume continues to consume the source topic where it left off, and the sinks continue draining from the Kafka channels where they left off. Post upgrade, remove the following deprecated properties from `flume.conf` because they are no longer used in CDH 5.8.0 or higher:

```
tier1.sources.source1.zookeeperConnect = zkhost:2181
tier1.sources.source1.topic = flumetopic
tier1.sources.source1.groupId = flume

tier1.channels.channel1.zookeeperConnect=zkhost:2181
tier1.channels.channel1.topic=flumechannel1
tier1.channels.channel1.groupId = flumechannel
```

Upgrading CDH 4 to CDH 5

This topic describes how to upgrade from CDH 4 to CDH 5. You can upgrade to CDH 5 in the Cloudera Manager Admin Console using parcels or packages. Using parcels vastly simplifies the upgrade process.



Important:

- If you upgrade using packages, all future upgrades must be done manually.
- You cannot perform a rolling upgrade from CDH 4 to CDH 5. There are incompatibilities between the major versions, so a rolling restart is not possible. Rolling upgrade is also *not* supported from CDH 5 Beta 2 to CDH 5.
- If you have just upgraded to Cloudera Manager 5, you must hard restart the Cloudera Manager Agents as described in the [\(Optional\) Deploy a Cloudera Manager Agent Upgrade](#) on page 49.
- **HBase** - After you upgrade you must recompile all HBase coprocessor and custom JARs.
- **Impala**
 - If you upgrade to CDH 5.1, Impala will be upgraded to 1.4.1. See [New Features in Impala](#) for information about Impala 1.4.x features.
 - If you upgrade to CDH 5.0, Impala will be upgraded to 1.3.2. If you have CDH 4 installed with Impala 1.4.0, Impala will be *downgraded* to Impala 1.3.2. See [New Features in Impala](#) for information about Impala 1.3 features.
- **Hive and Parquet**
 - When upgrading from CDH 4 to CDH 5, upgrade scripts may modify your schemas. For example, if you used Parquet with CDH 4, CDH 5 changes the input and output formats. The upgrade script changes `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` to `org.apache.hadoop.hive ql.io.parquet.MapredParquetInputFormat` and `org.apache.hadoop.hive ql.io.parquet.MapredParquetOutputFormat` in the schema. This may cause errors such as `Table already exists but schema doesn't match` and you may need to modify MapReduce jobs to use the newer Parquet Serdes.
- **MapReduce and YARN**
 - In a Cloudera Manager deployment of a CDH 5 cluster, the YARN service is the default MapReduce computation framework. In CDH 5, the MapReduce service has been deprecated. However, the MapReduce service is fully supported for backward compatibility through the CDH 5 lifecycle.
 - In a Cloudera Manager deployment of a CDH 4 cluster, the MapReduce service is the default MapReduce computation framework. You can create a YARN service in a CDH 4 cluster, but it is not considered production ready.
 - For production uses, Cloudera recommends that *only one* MapReduce framework should be running at any given time. If development needs or other use case requires switching between MapReduce and YARN, both services can be configured at the same time, but only one should be running (to fully optimize the hardware resources available).

For information on migrating from MapReduce to YARN, see [Managing YARN \(MRv2\) and MapReduce \(MRv1\)](#).



Warning: You can use Cloudera Manager to roll back an upgrade from CDH 4 to CDH 5 as long as you backup certain configuration files, databases, and other artifacts before beginning an upgrade. However, after you have [finalized the HDFS upgrade](#) you can no longer roll back the CDH upgrade. See [Rolling Back a CDH 4-to-CDH 5 Upgrade](#) on page 117 for the backup and rollback procedures.

Before You Begin

- Read the [CDH 5 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- [Upgrade to Cloudera Manager 5](#) before upgrading to CDH 5.

- Ensure Java 1.7 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130, and read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Make sure no Oozie workflows are in RUNNING or SUSPENDED status; otherwise, the Oozie database upgrade will fail and you must reinstall CDH 4 to complete or kill those running workflows.
- Delete Symbolic Links in HDFS

If there are symbolic links in HDFS when you upgrade from CDH 4 to CDH 5, the upgrade will fail and you must downgrade to CDH 4, delete the symbolic links, and start over.

1. Go to the directory on the NameNode that contains the latest `fsimage`. The location of this directory is specified as the value of `dfs.namenode.name.dir` (or `dfs.name.dir`) in `hdfs-site.xml`.
2. Use a command such as the following to write out the path names in the `fsimage`:

```
$ hdfs oiv -i FSIMAGE -o /tmp/YYYY-MM-DD_FSIMAGE.txt
```

3. Use a command such as the following to find the path names of any symbolic links listed in `/tmp/YYYY-MM-DD_FSIMAGE.txt` and write them out to the file `/tmp/symlinks.txt`:

```
$ grep -- "->" /tmp/YYYY-MM-DD_FSIMAGE.txt > /tmp/symlinks.txt
```

4. Delete any symbolic links listed in `/tmp/symlinks.txt`.

- When upgrading from CDH 4 to CDH 5, Oozie upgrade can take a very long time. For upgrades from CDH 4.3 and higher, you can reduce this time by reducing the amount of history Oozie retains:

1. Go to the Oozie service.
2. Click the **Configuration** tab.
3. Click **Category > Advanced**.
4. In **Oozie Server Advanced Configuration Snippet (Safety Valve) for oozie-site.xml**, enter the following

```
<property>
<name>oozie.service.PurgeService.older.than</name>
<value>7</value>
</property>
<property>
<name>oozie.service.PurgeService.purge.limit</name>
<value>1000</value>
</property>
```

5. For CDH lower than 5.2, enable DEBUG level logging:

- a. Click **Category > Logs**.
- b. Set **Oozie Server Logging Threshold** to **DEBUG**.

6. Click **Save Changes** to commit the changes.
7. Restart the Oozie Server role.

8. Wait for the purge service to run and finish. By default, the service runs every hour. The purge service emits the following messages in the Oozie server log:

```
STARTED Purge to purge Workflow Jobs older than [7] days, Coordinator Jobs older than [7] days, and Bundlejobs older than [7] days.  
ENDED Purge deleted [x] workflows, [y] coordinatorActions, [z] coordinators, [w] bundles
```

9. Revert the purge service and log level settings to the default.

- When upgrading from CDH 4 to CDH 5, Hue upgrade can take a very long time if the `beeswax_queryhistory`, `beeswax_savedquery`, and `oozie_job` tables are larger than 1000 records. You can reduce the upgrade time by running a script to reduce the size of the Hue database:

1. Stop the Hue service.
2. Back up the Hue database.
3. Download the [history cleanup script](#) to the host running the Hue Server.
4. Run the following as root:

- **parcel installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l  
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"  
/opt/cloudera/parcels/CDH/share/hue/build/env/bin/hue shell
```

- **package installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l  
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"  
/usr/share/hue/build/env/bin/hue shell
```

5. Run the downloaded script in the Hue shell.

- **If Using MySQL as Hue Backend:** You may face issues after the upgrade if the default engine for MySQL does not match the engine used by the Hue tables. To confirm the match:

1. Open the `my.cnf` file for MySQL, search for "default-storage-engine" and note its value.
2. Connect to MySQL and run the following commands:

```
use hue;  
show create table auth_user;
```

3. Search for the "ENGINE=" line and confirm that its value matches the one for the "default-storage-engine" above.

If the default engines do not match, Hue will display a warning on its start-up page ([http://\\$HUE_HOST:\\$HUE_PORT/about](http://$HUE_HOST:$HUE_PORT/about)). Work with your database administrator to convert the current Hue MySQL tables to the engine in use by MySQL, as noted by the "default-storage-engine" property.

- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.
- If using HBase:
 - Run `hbase hbck`.
 - Before you can upgrade HBase from CDH 4 to CDH 5, your HFiles must be upgraded from HFile v1 format to HFile v2, because CDH 5 no longer supports HFile v1. The upgrade procedure itself is different if you are using Cloudera Manager or the command line, but has the same results. The first step is to check for instances of

HFile v1 in the HFiles and mark them to be upgraded to HFile v2, and to check for and report about corrupted files or files with unknown versions, which need to be removed manually. The next step is to rewrite the HFiles during the next major compaction. After the HFiles are upgraded, you can continue the upgrade. After the upgrade is complete, you must recompile custom coprocessors and JARs. To check and upgrade the files:

1. In the Cloudera Admin Console, go to the HBase service and run **Actions > Check HFile Version**.
2. Check the output of the command in the `stderr` log.

Your output should be similar to the following:

```
Tables Processed:
hdfs://localhost:41020/myHBase/.META.
hdfs://localhost:41020/myHBase/usertable
hdfs://localhost:41020/myHBase/TestTable
hdfs://localhost:41020/myHBase/t

Count of HFileV1: 2
HFileV1:
hdfs://localhost:41020/myHBase/usertable
/fa02dac1f38d03577bd0f7e666f12812/family/249450144068442524
hdfs://localhost:41020/myHBase/usertable
/ecdd3eae2d2fcf8184ac025555bb2af/family/249450144068442512

Count of corrupted files: 1
Corrupted Files:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812/family/1
Count of Regions with HFileV1: 2
Regions to Major Compact:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812
hdfs://localhost:41020/myHBase/usertable/ecdd3eae2d2fcf8184ac025555bb2af
```

In the example above, you can see that the script has detected two HFile v1 files, one corrupt file and the regions to major compact.

3. Trigger a major compaction on each of the reported regions. This major compaction rewrites the files from HFile v1 to HFile v2 format. To run the major compaction, start HBase Shell and issue the `major_compact` command.

```
$ /usr/lib/hbase/bin/hbase shell
hbase> major_compact 'usertable'
```

You can also do this in a single step by using the `echo` shell built-in command.

```
$ echo "major_compact 'usertable'" | /usr/lib/hbase/bin/hbase shell
```

- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reen able Cloudera Manager alerts.

Upgrading from CDH 4 to CDH 5 Parcels

Minimum Required Role: [Full Administrator](#)

This topic covers upgrading a CDH 4 cluster to a CDH 5 cluster using the upgrade wizard, which will install CDH 5 parcels. Your CDH 4 cluster can be using *either parcels or packages*; you can use the cluster upgrade wizard to upgrade using parcels in either case.

If you want to upgrade using CDH 5 packages, you can do so using a manual process. See [Upgrading from CDH 4 Packages to CDH 5 Packages](#) on page 95.

The steps to upgrade a CDH installation managed by Cloudera Manager using parcels are as follows.

Before You Begin

- Read the [CDH 5 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- [Upgrade to Cloudera Manager 5](#) before upgrading to CDH 5.
- Ensure Java 1.7 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130, and read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Make sure no Oozie workflows are in RUNNING or SUSPENDED status; otherwise, the Oozie database upgrade will fail and you must reinstall CDH 4 to complete or kill those running workflows.
- Delete Symbolic Links in HDFS

If there are symbolic links in HDFS when you upgrade from CDH 4 to CDH 5, the upgrade will fail and you must downgrade to CDH 4, delete the symbolic links, and start over.

1. Go to the directory on the NameNode that contains the latest `fsimage`. The location of this directory is specified as the value of `dfs.namenode.name.dir` (or `dfs.name.dir`) in `hdfs-site.xml`.
2. Use a command such as the following to write out the path names in the `fsimage`:

```
$ hdfs oiv -i FSIMAGE -o /tmp/YYYY-MM-DD_FSIMAGE.txt
```

3. Use a command such as the following to find the path names of any symbolic links listed in `/tmp/YYYY-MM-DD_FSIMAGE.txt` and write them out to the file `/tmp/symlinks.txt`:

```
$ grep -- "->" /tmp/YYYY-MM-DD_FSIMAGE.txt > /tmp/symlinks.txt
```

4. Delete any symbolic links listed in `/tmp/symlinks.txt`.

- When upgrading from CDH 4 to CDH 5, Oozie upgrade can take a very long time. For upgrades from CDH 4.3 and higher, you can reduce this time by reducing the amount of history Oozie retains:

1. Go to the Oozie service.
2. Click the **Configuration** tab.
3. Click **Category > Advanced**.
4. In **Oozie Server Advanced Configuration Snippet (Safety Valve) for oozie-site.xml**, enter the following

```
<property>
<name>oozie.service.PurgeService.older.than</name>
<value>7</value>
</property>
<property>
<name>oozie.service.PurgeService.purge.limit</name>
<value>1000</value>
</property>
```

5. For CDH lower than 5.2, enable DEBUG level logging:

- a. Click **Category > Logs**.
- b. Set **Oozie Server Logging Threshold** to **DEBUG**.

6. Click **Save Changes** to commit the changes.
7. Restart the Oozie Server role.
8. Wait for the purge service to run and finish. By default, the service runs every hour. The purge service emits the following messages in the Oozie server log:

```
STARTED Purge to purge Workflow Jobs older than [7] days, Coordinator Jobs older than
[7] days, and Bundlejobs older than [7] days.
ENDED Purge deleted [x] workflows, [y] coordinatorActions, [z] coordinators, [w] bundles
```

9. Revert the purge service and log level settings to the default.

- When upgrading from CDH 4 to CDH 5, Hue upgrade can take a very long time if the `beeswax_queryhistory`, `beeswax_savedquery`, and `oozie_job` tables are larger than 1000 records. You can reduce the upgrade time by running a script to reduce the size of the Hue database:

1. Stop the Hue service.
2. Back up the Hue database.
3. Download the [history cleanup script](#) to the host running the Hue Server.
4. Run the following as root:

- **parcel installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"
/opt/cloudera/parcels/CDH/share/hue/build/env/bin/hue shell
```

- **package installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"
/usr/share/hue/build/env/bin/hue shell
```

5. Run the downloaded script in the Hue shell.

- **If Using MySQL as Hue Backend:** You may face issues after the upgrade if the default engine for MySQL does not match the engine used by the Hue tables. To confirm the match:

1. Open the `my.cnf` file for MySQL, search for "default-storage-engine" and note its value.
2. Connect to MySQL and run the following commands:

```
use hue;
show create table auth_user;
```

3. Search for the "ENGINE=" line and confirm that its value matches the one for the "default-storage-engine" above.

If the default engines do not match, Hue will display a warning on its start-up page (`http://$HUE_HOST:$HUE_PORT/about`). Work with your database administrator to convert the current Hue MySQL tables to the engine in use by MySQL, as noted by the "default-storage-engine" property.

- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.

- If using HBase:
 - Run `hbase hbck`.
 - Before you can upgrade HBase from CDH 4 to CDH 5, your HFiles must be upgraded from HFile v1 format to HFile v2, because CDH 5 no longer supports HFile v1. The upgrade procedure itself is different if you are using Cloudera Manager or the command line, but has the same results. The first step is to check for instances of HFile v1 in the HFiles and mark them to be upgraded to HFile v2, and to check for and report about corrupted files or files with unknown versions, which need to be removed manually. The next step is to rewrite the HFiles during the next major compaction. After the HFiles are upgraded, you can continue the upgrade. After the upgrade is complete, you must recompile custom coprocessors and JARs. To check and upgrade the files:
 1. In the Cloudera Admin Console, go to the HBase service and run **Actions > Check HFile Version**.
 2. Check the output of the command in the `stderr` log.

Your output should be similar to the following:

```
Tables Processed:
hdfs://localhost:41020/myHBase/.META.
hdfs://localhost:41020/myHBase/usertable
hdfs://localhost:41020/myHBase/TestTable
hdfs://localhost:41020/myHBase/t

Count of HFileV1: 2
HFileV1:
hdfs://localhost:41020/myHBase/usertable
/fa02dac1f38d03577bd0f7e666f12812/family/249450144068442524
hdfs://localhost:41020/myHBase/usertable
/ecdd3eaae2d2fcf8184ac025555bb2af/family/249450144068442512

Count of corrupted files: 1
Corrupted Files:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812/family/1
Count of Regions with HFileV1: 2
Regions to Major Compact:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812
hdfs://localhost:41020/myHBase/usertable/ecdd3eaae2d2fcf8184ac025555bb2af
```

In the example above, you can see that the script has detected two HFile v1 files, one corrupt file and the regions to major compact.

3. Trigger a major compaction on each of the reported regions. This major compaction rewrites the files from HFile v1 to HFile v2 format. To run the major compaction, start HBase Shell and issue the `major_compact` command.

```
$ /usr/lib/hbase/bin/hbase shell
hbase> major_compact 'usertable'
```

You can also do this in a single step by using the `echo` shell built-in command.

```
$ echo "major_compact 'usertable'" | /usr/lib/hbase/bin/hbase shell
```

- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reen able Cloudera Manager alerts.

Stop All Services

1. Stop the cluster.

- a. On the **Home > Status** tab, click



to the right of the cluster name and select **Stop**.

- b. Click **Stop** in the confirmation screen. The **Command Details** window shows the progress of stopping services.

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

2. Stop the Cloudera Management Service:

- a. Do one of the following:

- 1. Select **Clusters > Cloudera Management Service**.
- 2. Select **Actions > Stop**.

- 1. On the **Home > Status** tab, click



to the right of **Cloudera Management Service** and select **Stop**.

- b. Click **Stop** to confirm. The **Command Details** window shows the progress of stopping the roles.

- c. When **Command completed with n/n successful subcommands** appears, the task is complete. Click **Close**.

Perform Service-Specific Prerequisite Actions

- **Accumulo** - if you have installed the Accumulo parcel, deactivate it following the instructions in [Managing Parcels](#).
- **HDFS** - Back up HDFS metadata on the NameNode:

1. Go to the HDFS service.
2. Click the **Configuration** tab.
3. In the Search field, search for "NameNode Data Directories" and note the value.
4. On the active NameNode host, back up the directory listed in the NameNode Data Directories property. If more than one is listed, make a backup of one directory, because each directory is a complete copy. For example, if the NameNode data directory is `/data/dfs/nn`, do the following as root:

```
# cd /data/dfs/nn
# tar -cvf /root/nn_backup_data.tar .
```

You should see output like this:

```
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
./image/
./image/fsimage
```

If a file with the extension `lock` exists in the NameNode data directory, the NameNode most likely is still running. Repeat the steps, beginning with shutting down the NameNode role.

- Back up the Hive and Sqoop metastore databases.

1. For each affected service:

- a. If not already stopped, stop the service.

- b. Back up the database. See [Backing Up Databases](#).

Remove CDH Packages

If your previous installation of CDH was done using *packages*, you must remove those packages on all hosts in the cluster being upgraded. This will definitely be the case if you are running a version of CDH prior to CDH 4.1.3, since parcels were not available with those releases.

1. If your Hue service uses the embedded SQLite database, back up `/var/lib/hue/desktop.db` to a location that is not `/var/lib/hue` because this directory is removed when the packages are removed.
2. Uninstall the CDH packages. On each host:

Operating System	Command
RHEL	<code>\$ sudo yum remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc</code>
SLES	<code>\$ sudo zypper remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc</code>
Ubuntu or Debian	<code>\$ sudo apt-get purge bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc</code>

3. Restart all the Cloudera Manager Agents to force an update of the installed binaries reported by the Agent. On each host:


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

4. Run the Host Inspector to verify that the packages have been removed:
 - a. Click **Hosts** tab and then click the **Host Inspector** button.
 - b. When the command completes, click **Show Inspector Results**.
5. If your Hue service uses the embedded SQLite database, restore the database you backed up:
 - a. Stop the Hue service.
 - b. Copy the backup from the temporary location to the newly created Hue database directory, `/var/lib/hue`.
 - c. Start the Hue service.

Deactivate and Remove the GPL Extras Parcel

If you are using LZO, deactivate and remove the CDH 4 GPL Extras parcel.

Run the Upgrade Wizard

1. Log into the Cloudera Manager Admin console.
2. From the **Home > Status** tab, click  next to the cluster name and select **Upgrade Cluster**. The Upgrade Wizard starts.
3. If the option to choose between packages and parcels displays, select the **Use Parcels** option.
4. In the **Choose CDH Version (Parcels)** field, select the CDH version. If there are no qualifying parcels, click the **Modify the Remote Parcel Repository URLs** link to go to the [Parcel Configuration Settings](#) page where you can add the locations of parcel repositories. Click **Continue**.
5. Read the notices for steps you must complete before upgrading, click the **Yes, I ...** checkboxes after completing the steps, and click **Continue**.
6. Cloudera Manager checks that hosts have the correct software installed. Click **Continue**.
7. The selected parcels are downloaded and distributed. Click **Continue**.
8. The Host Inspector runs and displays the CDH version on the hosts. Click **Continue**.

9. Click **Continue**. Cloudera Manager performs all service upgrades and restarts the cluster.

10 The wizard reports the result of the upgrade. Choose one of the following:

- Leave **OK, set up YARN and import existing configuration from my MapReduce service** checked.
 1. Click **Continue** to proceed. Cloudera Manager stops the YARN service (if running) and its dependencies.
 2. Click **Continue** to proceed. The next page indicates some additional configuration required by YARN.
 3. Verify or modify the configurations and click **Continue**. The Switch Cluster to MR2 step proceeds.
 4. When all steps have completed, click **Continue**.
- Clear **OK, set up YARN and import existing configuration from my MapReduce service**.

11 Click **Finish** to return to the **Home > Status** tab.

12 (Optional) Remove the MapReduce service.

a. In the MapReduce row, right-click



and select **Delete**. Click **Delete** to confirm.

Recover from Failed Steps



Note: If you encounter errors during these steps:

- If the converting configuration parameters step fails, Cloudera Manager rolls back all configurations to CDH 4. Fix any reported problems and retry the upgrade.
- If the upgrade command fails at any point after the convert configuration step, there is no retry support in Cloudera Manager. You must first correct the error, then manually re-run the individual commands. You can view the remaining commands in the Recent Commands page.
- If the HDFS upgrade metadata step fails, you cannot revert back to CDH 4 unless you restore a backup of Cloudera Manager.

The actions performed by the upgrade wizard are listed in [Upgrade Wizard Actions](#) on page 94. If any of the steps in the **Command Progress** screen fails, complete the step as described in that section before proceeding.

Upgrade the GPL Extras Parcel

If you are using LZO:

1. Install the CDH 5 GPL Extras parcel. See [Installing the GPL Extras Parcel](#).
2. Reconfigure and restart services that use the parcel. See [Configuring Services to Use the GPL Extras Parcel](#).

Restart the Reports Manager Role

1. Do one of the following:
 - Select **Clusters > Cloudera Management Service**.
 - On the **Home > Status** tab, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.
2. Click the **Instances** tab.
3. Check the checkbox next to **Reports Manager**.
4. Select **Actions for Selected > Restart** and then **Restart** to confirm.

Recompile JARs

- **MapReduce and YARN** - Recompile JARs used in MapReduce applications. For further information, see [For MapReduce Programmers: Writing and Running Jobs](#).

Upgrading CDH and Managed Services Using Cloudera Manager

- **HBase** - Recompile coprocessor and custom JARs used by HBase applications.

Finalize the HDFS Metadata Upgrade

Finalize the HDFS metadata upgrade. To determine when finalization is warranted, run important workloads and ensure they are successful.

1. Go to the HDFS service.
2. Click the **Instances** tab.
3. Click the **NameNode** instance.
4. Select **Actions > Finalize Metadata Upgrade** and click **Finalize Metadata Upgrade** to confirm.

Upgrade Wizard Actions

Do the steps in this section only if the upgrade wizard reports a failure.

Upgrade HDFS Metadata

1. Start the ZooKeeper service.
2. Go to the HDFS service.
3. Select **Actions > Upgrade HDFS Metadata** and click **Upgrade HDFS Metadata** to confirm.

Upgrade HBase

1. Go to the HBase service.
2. Select **Actions > Upgrade HBase** and click **Upgrade HBase** to confirm.

Upgrade the Hive Metastore Database

1. Go to the Hive service.
2. Select **Actions > Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
3. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade Oozie

1. Go to the Oozie service.
2. Select **Actions > Upgrade Database** and click **Upgrade Database** to confirm.
3. Start the Oozie service.
4. Select **Actions > Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.
2. Select **Actions > Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Start Cluster Services

1. On the **Home > Status** tab, click




to the right of the cluster name and select **Start**.

2. Click **Start** that appears in the next screen to confirm. The **Command Details** window shows the progress of starting services.

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

Deploy Client Configuration Files

1. On the Home page, click  to the right of the cluster name and select **Deploy Client Configuration**.
2. Click the **Deploy Client Configuration** button in the confirmation pop-up that appears.

Upgrading from CDH 4 Packages to CDH 5 Packages

Minimum Required Role: [Full Administrator](#)

If you originally used Cloudera Manager to install CDH using packages, you can upgrade to CDH 5 either using packages or parcels. Parcels is the preferred and recommended way to upgrade, as the upgrade wizard provided for parcels handles the upgrade process almost completely automatically.

The steps to upgrade a CDH installation managed by Cloudera Manager using packages are as follows.

Before You Begin

- Read the [CDH 5 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- [Upgrade to Cloudera Manager 5](#) before upgrading to CDH 5.
- Ensure Java 1.7 is installed across the cluster. For installation instructions and recommendations, see [Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment](#) on page 130, and read [Known Issues and Workarounds in Cloudera Manager 5](#) before you proceed with the upgrade.
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Make sure no Oozie workflows are in RUNNING or SUSPENDED status; otherwise, the Oozie database upgrade will fail and you must reinstall CDH 4 to complete or kill those running workflows.
- Delete Symbolic Links in HDFS

If there are symbolic links in HDFS when you upgrade from CDH 4 to CDH 5, the upgrade will fail and you must downgrade to CDH 4, delete the symbolic links, and start over.

1. Go to the directory on the NameNode that contains the latest `fsimage`. The location of this directory is specified as the value of `dfs.namenode.name.dir` (or `dfs.name.dir`) in `hdfs-site.xml`.
2. Use a command such as the following to write out the path names in the `fsimage`:

```
$ hdfs oiv -i FSIMAGE -o /tmp/YYYY-MM-DD_FSIMAGE.txt
```

3. Use a command such as the following to find the path names of any symbolic links listed in `/tmp/YYYY-MM-DD_FSIMAGE.txt` and write them out to the file `/tmp/symlinks.txt`:

```
$ grep -- "->" /tmp/YYYY-MM-DD_FSIMAGE.txt > /tmp/symlinks.txt
```

4. Delete any symbolic links listed in `/tmp/symlinks.txt`.
- When upgrading from CDH 4 to CDH 5, Oozie upgrade can take a very long time. For upgrades from CDH 4.3 and higher, you can reduce this time by reducing the amount of history Oozie retains:
 1. Go to the Oozie service.
 2. Click the **Configuration** tab.
 3. Click **Category > Advanced**.

4. In **Oozie Server Advanced Configuration Snippet (Safety Valve)** for `oozie-site.xml`, enter the following

```
<property>
<name>oozie.service.PurgeService.older.than</name>
<value>7</value>
</property>
<property>
<name>oozie.service.PurgeService.purge.limit</name>
<value>1000</value>
</property>
```

5. For CDH lower than 5.2, enable DEBUG level logging:

- a. Click **Category > Logs**.
- b. Set **Oozie Server Logging Threshold** to **DEBUG**.

6. Click **Save Changes** to commit the changes.

7. Restart the Oozie Server role.

8. Wait for the purge service to run and finish. By default, the service runs every hour. The purge service emits the following messages in the Oozie server log:

```
STARTED Purge to purge Workflow Jobs older than [7] days, Coordinator Jobs older than [7] days, and Bundlejobs older than [7] days.
ENDED Purge deleted [x] workflows, [y] coordinatorActions, [z] coordinators, [w] bundles
```

9. Revert the purge service and log level settings to the default.

- When upgrading from CDH 4 to CDH 5, Hue upgrade can take a very long time if the `beeswax_queryhistory`, `beeswax_savedquery`, and `oozie_job` tables are larger than 1000 records. You can reduce the upgrade time by running a script to reduce the size of the Hue database:

1. Stop the Hue service.
2. Back up the Hue database.
3. Download the [history cleanup script](#) to the host running the Hue Server.
4. Run the following as root:

• **parcel installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l /var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"
/opt/cloudera/parcels/CDH/share/hue/build/env/bin/hue shell
```

• **package installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l /var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"
/usr/share/hue/build/env/bin/hue shell
```

5. Run the downloaded script in the Hue shell.

- **If Using MySQL as Hue Backend:** You may face issues after the upgrade if the default engine for MySQL does not match the engine used by the Hue tables. To confirm the match:

1. Open the `my.cnf` file for MySQL, search for `"default-storage-engine"` and note its value.
2. Connect to MySQL and run the following commands:

```
use hue;
show create table auth_user;
```

3. Search for the `"ENGINE="` line and confirm that its value matches the one for the `"default-storage-engine"` above.

If the default engines do not match, Hue will display a warning on its start-up page ([http://\\$HUE_HOST:\\$HUE_PORT/about](http://$HUE_HOST:$HUE_PORT/about)). Work with your database administrator to convert the current Hue MySQL tables to the engine in use by MySQL, as noted by the "default-storage-engine" property.

- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.
- If using HBase:
 - Run `hbase hbck`.
 - Before you can upgrade HBase from CDH 4 to CDH 5, your HFiles must be upgraded from HFile v1 format to HFile v2, because CDH 5 no longer supports HFile v1. The upgrade procedure itself is different if you are using Cloudera Manager or the command line, but has the same results. The first step is to check for instances of HFile v1 in the HFiles and mark them to be upgraded to HFile v2, and to check for and report about corrupted files or files with unknown versions, which need to be removed manually. The next step is to rewrite the HFiles during the next major compaction. After the HFiles are upgraded, you can continue the upgrade. After the upgrade is complete, you must recompile custom coprocessors and JARs. To check and upgrade the files:
 1. In the Cloudera Admin Console, go to the HBase service and run **Actions > Check HFile Version**.
 2. Check the output of the command in the `stderr` log.

Your output should be similar to the following:

```
Tables Processed:
hdfs://localhost:41020/myHBase/.META.
hdfs://localhost:41020/myHBase/usertable
hdfs://localhost:41020/myHBase/TestTable
hdfs://localhost:41020/myHBase/t

Count of HFileV1: 2
HFileV1:
hdfs://localhost:41020/myHBase/usertable
/fa02dac1f38d03577bd0f7e666f12812/family/249450144068442524
hdfs://localhost:41020/myHBase/usertable
/ecdd3eaae2d2fcf8184ac025555bb2af/family/249450144068442512

Count of corrupted files: 1
Corrupted Files:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812/family/1
Count of Regions with HFileV1: 2
Regions to Major Compact:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812
hdfs://localhost:41020/myHBase/usertable/ecdd3eaae2d2fcf8184ac025555bb2af
```

In the example above, you can see that the script has detected two HFile v1 files, one corrupt file and the regions to major compact.

3. Trigger a major compaction on each of the reported regions. This major compaction rewrites the files from HFile v1 to HFile v2 format. To run the major compaction, start HBase Shell and issue the `major_compact` command.

```
$ /usr/lib/hbase/bin/hbase shell
hbase> major_compact 'usertable'
```

You can also do this in a single step by using the `echo` shell built-in command.

```
$ echo "major_compact 'usertable'" | /usr/lib/hbase/bin/hbase shell
```

- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.

Stop All Services

1. Stop the cluster.

- a. On the **Home > Status** tab, click



to the right of the cluster name and select **Stop**.

- b. Click **Stop** in the confirmation screen. The **Command Details** window shows the progress of stopping services.

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

2. Stop the Cloudera Management Service:

- a. Do one of the following:

- 1. Select **Clusters > Cloudera Management Service**.
- 2. Select **Actions > Stop**.

- 1. On the **Home > Status** tab, click



to the right of **Cloudera Management Service** and select **Stop**.

- b. Click **Stop** to confirm. The **Command Details** window shows the progress of stopping the roles.

- c. When **Command completed with n/n successful subcommands** appears, the task is complete. Click **Close**.

Perform Service-Specific Prerequisite Actions

- **HDFS** - Back up HDFS metadata on the NameNode:

1. Go to the HDFS service.
2. Click the **Configuration** tab.
3. In the Search field, search for "NameNode Data Directories" and note the value.
4. On the active NameNode host, back up the directory listed in the NameNode Data Directories property. If more than one is listed, make a backup of one directory, because each directory is a complete copy. For example, if the NameNode data directory is `/data/dfs/nn`, do the following as root:

```
# cd /data/dfs/nn
# tar -cvf /root/nn_backup_data.tar .
```

You should see output like this:

```
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
```

```
./image/
./image/fsimage
```

If a file with the extension *lock* exists in the NameNode data directory, the NameNode most likely is still running. Repeat the steps, beginning with shutting down the NameNode role.

- Back up the Hive and Sqoop metastore databases.
 1. For each affected service:
 - a. If not already stopped, stop the service.
 - b. Back up the database. See [Backing Up Databases](#).

Uninstall CDH 4

Uninstall CDH 4 on each host as follows:

Operating System	Command
RHEL	\$ sudo yum remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc
SLES	\$ sudo zypper remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc
Ubuntu or Debian	\$ sudo apt-get purge bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc

Remove CDH 4 Repository Files

Remove all Cloudera CDH 4 repository files. For example, on an RHEL or similar system, remove all files in `/etc/yum.repos.d` that have `cloudera` as part of the name.



Important:

- Before removing the files, make sure you have not added any custom entries that you want to preserve. (To preserve custom entries, back up the files before removing them.)
- Make sure you remove Impala and Search repository files, as well as the CDH repository file.

Install CDH 5 Components

• Red Hat

1. Download and install the "1-click Install" package.
 - a. Download the CDH 5 "1-click Install" package (or RPM).

Click the appropriate RPM and **Save File** to a directory with write access (for example, your home directory).

OS Version	Link to CDH 5 RPM
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS/Oracle 6	RHEL/CentOS/Oracle 6 link
RHEL/CentOS/Oracle 7	RHEL/CentOS/Oracle 7 link

b. Install the RPM for all RHEL versions:

```
$ sudo yum --nogpgcheck localinstall cloudera-cdh-5-0.x86_64.rpm
```

2. (Optionally) add a repository key:

- **Red Hat/CentOS/Oracle 5**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **Red Hat/CentOS/Oracle 6**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

3. Install the CDH packages:

```
$ sudo yum clean all  
$ sudo yum install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-hdfs-nfs3  
hadoop-httpfs hadoop-kms hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase  
hue-impala hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper  
impala impala-shell kite llama mahout oozie pig pig-udf-datafu search sentry  
solr-mapreduce spark-core spark-master spark-worker spark-history-server spark-python  
sqoop sqoop2 whirr
```



Note: Installing these packages also installs all the other CDH packages required for a full CDH 5 installation.

- **SLES**

1. Download and install the "1-click Install" package.

a. Download the CDH 5 "1-click Install" package.

Download the [RPM file](#), choose **Save File**, and save it to a directory to which you have write access (for example, your home directory).

b. Install the RPM:

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

c. Update your system package index by running the following:

```
$ sudo zypper refresh
```

2. (Optionally) add a repository key:

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

3. Install the CDH packages:

```
$ sudo zypper clean --all  
$ sudo zypper install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-hdfs-nfs3  
hadoop-httpfs hadoop-kms hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase  
hue-impala hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper  
impala impala-shell kite llama mahout oozie pig pig-udf-datafu search sentry  
solr-mapreduce spark-core spark-master spark-worker spark-history-server spark-python  
sqoop sqoop2 whirr
```



Note: Installing these packages also installs all the other CDH packages required for a full CDH 5 installation.

- **Ubuntu and Debian**

1. Download and install the "1-click Install" package

- a. Download the CDH 5 "1-click Install" package:

OS Version	Package Link
Jessie	Jessie package
Wheezy	Wheezy package
Precise	Precise package
Trusty	Trusty package

- b. Install the package by doing one of the following:

- Choose **Open with** in the download window to use the package manager.
- Choose **Save File**, save the package to a directory to which you have write access (for example, your home directory), and install it from the command line. For example:

```
sudo dpkg -i cdh5-repository_1.0_all.deb
```

2. Optionally add a repository key:

- **Debian Wheezy**

```
$ curl -s https://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key | sudo apt-key add -
```

- **Ubuntu Precise**

```
$ curl -s https://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key | sudo apt-key add -
```

3. Install the CDH packages:



```
$ sudo apt-get update
$ sudo apt-get install avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-hdfs-nfs3
hadoop-httpfs hadoop-kms hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase
hue-impala hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper
impala impala-shell kite llama mahout oozie pig pig-udf-datafu search sentry
solr-mapreduce spark-core spark-master spark-worker spark-history-server spark-python
sqoop sqoop2 whirr
```



Note: Installing these packages also installs all the other CDH packages required for a full CDH 5 installation.

Run the Upgrade Wizard

1. Log into the Cloudera Manager Admin console.

2. From the **Home > Status** tab, click  next to the cluster name and select **Upgrade Cluster**. The Upgrade Wizard starts.
3. If the option to choose between packages and parcels displays, select the **Use Parcels** option.
4. In the **Choose CDH Version (Parcels)** field, select the CDH version. If there are no qualifying parcels, click the **Modify the Remote Parcel Repository URLs** link to go to the [Parcel Configuration Settings](#) page where you can add the locations of parcel repositories. Click **Continue**.
5. Read the notices for steps you must complete before upgrading, click the **Yes, I ...** checkboxes after completing the steps, and click **Continue**.
6. Cloudera Manager checks that hosts have the correct software installed. Click **Continue**.
7. The selected parcels are downloaded and distributed. Click **Continue**.
8. The Host Inspector runs and displays the CDH version on the hosts. Click **Continue**.
9. Click **Continue**. Cloudera Manager performs all service upgrades and restarts the cluster.
- 10 The wizard reports the result of the upgrade. Choose one of the following:
 - Leave **OK, set up YARN and import existing configuration from my MapReduce service** checked.
 1. Click **Continue** to proceed. Cloudera Manager stops the YARN service (if running) and its dependencies.
 2. Click **Continue** to proceed. The next page indicates some additional configuration required by YARN.
 3. Verify or modify the configurations and click **Continue**. The Switch Cluster to MR2 step proceeds.
 4. When all steps have completed, click **Continue**.
 - Clear **OK, set up YARN and import existing configuration from my MapReduce service**.
- 11 Click **Finish** to return to the **Home > Status** tab.
- 12 (Optional) Remove the MapReduce service.
 - a. In the MapReduce row, right-click  and select **Delete**. Click **Delete** to confirm.

Recover from Failed Steps



Note: If you encounter errors during these steps:

- If the converting configuration parameters step fails, Cloudera Manager rolls back all configurations to CDH 4. Fix any reported problems and retry the upgrade.
- If the upgrade command fails at any point after the convert configuration step, there is no retry support in Cloudera Manager. You must first correct the error, then manually re-run the individual commands. You can view the remaining commands in the Recent Commands page.
- If the HDFS upgrade metadata step fails, you cannot revert back to CDH 4 unless you restore a backup of Cloudera Manager.

The actions performed by the upgrade wizard are listed in [Upgrade Wizard Actions](#) on page 103. If any of the steps in the **Command Progress** screen fails, complete the step as described in that section before proceeding.

Restart the Reports Manager Role

1. Do one of the following:
 - Select **Clusters > Cloudera Management Service**.
 - On the **Home > Status** tab, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.
2. Click the **Instances** tab.
3. Check the checkbox next to **Reports Manager**.

4. Select **Actions for Selected** > **Restart** and then **Restart** to confirm.

Recompile JARs

- **MapReduce and YARN** - Recompile JARs used in MapReduce applications. For further information, see [For MapReduce Programmers: Writing and Running Jobs](#).
- **HBase** - Recompile coprocessor and custom JARs used by HBase applications.

Finalize the HDFS Metadata Upgrade

Finalize the HDFS metadata upgrade. To determine when finalization is warranted, run important workloads and ensure they are successful.

1. Go to the HDFS service.
2. Click the **Instances** tab.
3. Click the **NameNode** instance.
4. Select **Actions** > **Finalize Metadata Upgrade** and click **Finalize Metadata Upgrade** to confirm.

Upgrade Wizard Actions

Do the steps in this section only if the upgrade wizard reports a failure.

Upgrade HDFS Metadata

1. Start the ZooKeeper service.
2. Go to the HDFS service.
3. Select **Actions** > **Upgrade HDFS Metadata** and click **Upgrade HDFS Metadata** to confirm.

Upgrade HBase

1. Go to the HBase service.
2. Select **Actions** > **Upgrade HBase** and click **Upgrade HBase** to confirm.

Upgrade the Hive Metastore Database

1. Go to the Hive service.
2. Select **Actions** > **Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
3. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade Oozie

1. Go to the Oozie service.
2. Select **Actions** > **Upgrade Database** and click **Upgrade Database** to confirm.
3. Start the Oozie service.
4. Select **Actions** > **Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.
2. Select **Actions** > **Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Start Cluster Services

1. On the **Home** > **Status** tab, click




to the right of the cluster name and select **Start**.

Upgrading CDH and Managed Services Using Cloudera Manager

2. Click **Start** that appears in the next screen to confirm. The **Command Details** window shows the progress of starting services.

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

Deploy Client Configuration Files

1. On the Home page, click  to the right of the cluster name and select **Deploy Client Configuration**.
2. Click the **Deploy Client Configuration** button in the confirmation pop-up that appears.

Upgrading CDH 4

Use the instructions in this section to upgrade to a higher CDH 4 minor release, that is from CDH 4.*a.x* to CDH 4. *b.y*. For example, CDH 4.6.0 to CDH 4.7.1.

You can upgrade to CDH 4.1.3 (or higher) within the Cloudera Manager Admin Console, using parcels and an upgrade wizard. This vastly simplifies the upgrade process. Electing to upgrade using packages means that future upgrades will still need to be done manually. Upgrading to a CDH 4 release prior to CDH 4.1.3 is possible using packages, though upgrading to a more current release is strongly recommended.

If you use parcels, have a Cloudera Enterprise license, and have enabled HDFS high availability, you can perform a [rolling upgrade](#) that lets you avoid cluster downtime.



Important: The following instructions describe how to upgrade from a CDH 4 release to a newer CDH 4 release in a Cloudera Manager deployment. If you are running CDH 3, you must upgrade to CDH 4 using the instructions at [Upgrading CDH 3 to CDH 4 in a Cloudera Managed Deployment](#).

To upgrade from CDH 4 to CDH 5, see [Upgrading CDH 4 to CDH 5](#) on page 83.

Before You Begin

- Before upgrading, be sure to read about the latest Incompatible Changes and Known Issues and Workarounds in the [CDH 4 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.
- Run `hbase hbck`.
- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the

upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.

- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.

Upgrade Procedures



Important:

- **Impala** - If you have CDH 4.1.x with Cloudera Impala installed, and you plan to upgrade to CDH 4.2 or higher, you must also upgrade Impala to version 1.2.1 or higher. With a parcel installation you can download and activate both parcels before you proceed to restart the cluster. You will need to change the remote parcel repo URL to point to the location of the released product as described in the upgrade procedures referenced below.
- **HBase** - In CDH 4.1.x, an HBase table could have an owner that had full administrative permissions on the table. The owner construct was removed as of CDH 4.2.0, and the code now relies exclusively on entries in the ACL table. Since table owners do not have an entry in this table, their permissions are removed on upgrade from CDH 4.1.x to CDH 4.2.0 or higher. If you are upgrading from CDH 4.1.x to CDH 4.2 or higher, and using HBase, you must add permissions for HBase owner users to the HBase ACL table before you perform the upgrade. See the Known Issues in the CDH 4 Release Notes, specifically the item "Must explicitly add permissions for owner users before upgrading from 4.1.x" in the [Known Issues in Apache HBase](#) section.
- **Hive** - Hive has undergone major version changes from CDH 4.0 to 4.1 and between CDH 4.1 and 4.2. (CDH 4.0 had Hive 0.8.0, CDH 4.1 used Hive 0.9.0, and 4.2 or higher has 0.10.0). This requires you to manually back up and upgrade the Hive metastore database when upgrading between major Hive versions. If you are upgrading from a version of CDH 4 prior to CDH 4.2 to a newer CDH 4 version, you must follow the steps for upgrading the metastore included in the upgrade procedures referenced below.

Upgrading CDH 4 Using Parcels

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

You can upgrade your CDH 4 cluster to a higher minor version of CDH 4 using parcels from within the Cloudera Manager Admin Console. Your current CDH 4 cluster can have been installed with either parcels or packages. The new version will use parcels.

The following procedure requires cluster downtime. If you use parcels, have a Cloudera Enterprise license, and have enabled HDFS high availability, you can perform a [rolling upgrade](#) that lets you avoid cluster downtime.



Important:

- **Impala** - If you have CDH 4.1.x with Cloudera Impala installed, and you plan to upgrade to CDH 4.2 or higher, you must also upgrade Impala to version 1.2.1 or higher. With a parcel installation you can download and activate both parcels before you proceed to restart the cluster. You will need to change the remote parcel repo URL to point to the location of the released product as described in the upgrade procedures referenced below.
- **HBase** - In CDH 4.1.x, an HBase table could have an owner that had full administrative permissions on the table. The owner construct was removed as of CDH 4.2.0, and the code now relies exclusively on entries in the ACL table. Since table owners do not have an entry in this table, their permissions are removed on upgrade from CDH 4.1.x to CDH 4.2.0 or higher. If you are upgrading from CDH 4.1.x to CDH 4.2 or higher, and using HBase, you must add permissions for HBase owner users to the HBase ACL table before you perform the upgrade. See the Known Issues in the CDH 4 Release Notes, specifically the item "Must explicitly add permissions for owner users before upgrading from 4.1.x" in the [Known Issues in Apache HBase](#) section.
- **Hive** - Hive has undergone major version changes from CDH 4.0 to 4.1 and between CDH 4.1 and 4.2. (CDH 4.0 had Hive 0.8.0, CDH 4.1 used Hive 0.9.0, and 4.2 or higher has 0.10.0). This requires you to manually back up and upgrade the Hive metastore database when upgrading between major Hive versions. If you are upgrading from a version of CDH 4 prior to CDH 4.2 to a newer CDH 4 version, you must follow the steps for upgrading the metastore included in the upgrade procedures referenced below.

To upgrade your version of CDH using parcels, the steps are as follows.

Before You Begin



- Before upgrading, be sure to read about the latest Incompatible Changes and Known Issues and Workarounds in the [CDH 4 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.
- Run `hbase hbck`.
- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reen able Cloudera Manager alerts.

Download, Distribute, and Activate Parcels

1.

In the Cloudera Manager Admin Console, click the Parcels indicator in the top navigation bar ( or ) to go to the Parcels page.

2. In the parcels page, click **Download** for the version(s) you want to download. If the parcel you want is not shown here — for example, you want to upgrade to version of CDH that is not the most current version — you can make additional parcel repos available through the [parcel settings](#) page. If your Cloudera Manager server does not have Internet access, you can obtain the required parcel file(s) and put them into the local repository. See [Creating and Using a Parcel Repository for Cloudera Manager](#) for more details.
3. When the download has completed, click **Distribute** for the version you downloaded.
4. When the parcel has been distributed and unpacked, the button will change to say **Activate**.
5. Click **Activate**. You are asked if you want to restart the cluster. *Do not restart the cluster at this time.*
6. Click **Close**.

Upgrade the Hive Metastore Database

Required if you are upgrading from an earlier version of CDH 4 to CDH 4.2 or higher.

1. Go to the Hive service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
4. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade the Oozie ShareLib

1. Go to the Oozie service.
2. Select **Actions > Start** and click **Start** to confirm.
3. Select **Actions > Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Restart the Services

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Restart**.

2. Click the **Restart** button in the confirmation pop-up that appears. The **Command Details** window shows the progress of starting services.

Deploy Client Configuration Files

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Deploy Client Configuration**.

2. Click the **Deploy Client Configuration** button in the confirmation pop-up that appears.

Remove the Previous CDH Version Packages and Refresh Symlinks

If your previous installation of CDH was done using packages, remove those packages on all hosts on which you installed the parcels and refresh the symlinks so that clients will run the new software versions. *Skip this step if your previous installation was using parcels.*

1. If Hue is configured to use SQLite as its database:
 - a. Stop the Hue service.
 - b. Back up the `desktop.db` to a temporary location before deleting the old Hue Common package. The location of the database can be found in the Hue service **Configuration** tab under **Service > Database > Hue's Database Directory**.



Important: Removing the Hue Common package will remove your Hue database; if you do not back it up you may lose all your Hue user account information.

2. Uninstall the CDH packages on each host:

- **Not including Impala and Search**

Operating System	Command
RHEL	<code>\$ sudo yum remove hadoop hue-common bigtop-jsvc bigtop-tomcat</code>
SLES	<code>\$ sudo zypper remove hadoop hue-common bigtop-jsvc bigtop-tomcat</code>
Ubuntu or Debian	<code>\$ sudo apt-get purge hadoop hue-common bigtop-jsvc bigtop-tomcat</code>

- **Including Impala and Search**

Operating System	Command
RHEL	<code>\$ sudo yum remove hadoop hue-common impala-shell solr-server 'bigtop-*</code>
SLES	<code>\$ sudo zypper remove hadoop hue-common impala-shell solr-server 'bigtop-*</code>
Ubuntu or Debian	<code>\$ sudo apt-get purge hadoop hue-common impala-shell solr-server 'bigtop-*</code>

3. Restart all the Cloudera Manager Agents to force an update of the symlinks to point to the newly installed components on each host:

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

Restore Backed up Hue Database

Restore the backup you created in [Remove the Previous CDH Version Packages and Refresh Symlinks](#) on page 108.

1. Go to the Hue service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Copy the backup from the temporary location to the newly created Hue database directory:
`/opt/cloudera/parcels/CDH-4.x.0-x.cd4.x.0.p0.xx/share/hue/desktop.`
4. Restart the Hue service.

Upgrading CDH 4 Using Packages

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

If you originally used Cloudera Manager to install your CDH service using packages, you can upgrade to a higher minor version of CDH 4 either using packages or parcels. Parcels is the preferred and recommended way to upgrade, as the upgrade wizard provided for parcels handles the upgrade process almost completely automatically.


Important:

- **Impala** - If you have CDH 4.1.x with Cloudera Impala installed, and you plan to upgrade to CDH 4.2 or higher, you must also upgrade Impala to version 1.2.1 or higher. With a parcel installation you can download and activate both parcels before you proceed to restart the cluster. You will need to change the remote parcel repo URL to point to the location of the released product as described in the upgrade procedures referenced below.
- **HBase** - In CDH 4.1.x, an HBase table could have an owner that had full administrative permissions on the table. The owner construct was removed as of CDH 4.2.0, and the code now relies exclusively on entries in the ACL table. Since table owners do not have an entry in this table, their permissions are removed on upgrade from CDH 4.1.x to CDH 4.2.0 or higher. If you are upgrading from CDH 4.1.x to CDH 4.2 or higher, and using HBase, you must add permissions for HBase owner users to the HBase ACL table before you perform the upgrade. See the Known Issues in the CDH 4 Release Notes, specifically the item "Must explicitly add permissions for owner users before upgrading from 4.1.x" in the [Known Issues in Apache HBase](#) section.
- **Hive** - Hive has undergone major version changes from CDH 4.0 to 4.1 and between CDH 4.1 and 4.2. (CDH 4.0 had Hive 0.8.0, CDH 4.1 used Hive 0.9.0, and 4.2 or higher has 0.10.0). This requires you to manually back up and upgrade the Hive metastore database when upgrading between major Hive versions. If you are upgrading from a version of CDH 4 prior to CDH 4.2 to a newer CDH 4 version, you must follow the steps for upgrading the metastore included in the upgrade procedures referenced below.

To upgrade your version of CDH using packages, the steps are as follows.

Before You Begin

- Before upgrading, be sure to read about the latest Incompatible Changes and Known Issues and Workarounds in the [CDH 4 Release Notes](#).
- Read the [Cloudera Manager 5 Release Notes](#).
- Ensure that the Cloudera Manager minor version is *equal to or greater than* the CDH minor version. For example:

Target CDH Version	Minimum Cloudera Manager Version
5.0.5	5.0.x
5.1.4	5.1.x
5.4.1	5.4.x

- Run the [Host Inspector](#) and fix every issue.
- If using security, run the [Running the Security Inspector](#).
- When upgrading Impala, whether in CDH or a standalone parcel or package, check your SQL against the newest reserved words listed in [incompatible changes](#). If upgrading across multiple versions or in case of any problems, check against the full list of [Impala keywords](#).
- Run `hdfs fsck /` and `hdfs dfsadmin -report` and fix every issue.
- Run `hbase hbck`.
- Review the upgrade procedure and reserve a maintenance window with enough time allotted to perform all steps. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks

and configuration validations. Exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.

Upgrade Unmanaged Components

Upgrading unmanaged components is a process that is separate from upgrading managed components. Upgrade the unmanaged components before proceeding to upgrade managed components. Components that you might have installed that are not managed by Cloudera Manager include:

- Pig
- Whirr
- Mahout

For information on upgrading these unmanaged components, see [CDH 4 Installation Guide](#).

Upgrade Managed Components

Use one of the following strategies to upgrade CDH 4:

- Use your operating system's package management tools to update all packages to the latest version using standard repositories. This approach works well because it minimizes the amount of configuration required and uses the simplest commands. Be aware that this can take a considerable amount of time if you have not upgraded the system recently. To update all packages on your system, use the following command:

Operating System	Command
RHEL	\$ sudo yum update
SLES	\$ sudo zypper up
Ubuntu or Debian	\$ sudo apt-get upgrade

- Use the `cloudera.com` repository that is added during a typical installation, only updating Cloudera components. This limits the scope of updates to be completed, so the process takes less time, however this process will not work if you created and used a custom repository. To install the new version, you can upgrade from the Cloudera repository by adding an entry to your operating system's package management configuration file. The repository location varies by operating system:

Operating System	Configuration File Repository Entry
Red Hat	https://archive.cloudera.com/cdh4/redhat/6/x86_64/cdh/4/
SLES	https://archive.cloudera.com/cdh4/sles/11/x86_64/cdh/4/
Debian Squeeze	[arch=amd64] https://archive.cloudera.com/cdh4/debian/squeeze squeeze-cdh4 contrib
Ubuntu Lucid	[arch=amd64] https://archive.cloudera.com/cdh4/ubuntu/lucid/amd64/cdh lucid-cdh4 contrib
Ubuntu Precise	[arch=amd64] https://archive.cloudera.com/cdh4/ubuntu/precise/amd64/cdh precise-cdh4 contrib

For example, under Red Hat, to upgrade from the Cloudera repository you can run commands such as the following on the CDH host to update only CDH:

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

**Note:**

- `cloudera-cdh4` is the name of the repository on your system; the name is usually in square brackets on the first line of the repo file, in this example `/etc/yum.repos.d/cloudera-cdh4.repo`:

```
[chris@ca727 yum.repos.d]$ more cloudera-cdh4.repo
[cloudera-cdh4]
...
```

- `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.

On a SLES system, use commands like this to clean cached repository information and then update only the CDH components. For example:

```
$ sudo zypper clean --all
$ sudo zypper up -r https://archive.cloudera.com/cdh4/sles/11/x86_64/cdh/4
```

To verify the URL, open the Cloudera repo file in `/etc/zypp/repos.d` on your system (for example `/etc/zypp/repos.d/cloudera-cdh4.repo`) and look at the line beginning

```
baseurl=
```

Use that URL in your `sudo zypper up -r` command.

On a Debian/Ubuntu system, use commands like this to clean cached repository information and then update only the CDH components. First:

```
$ sudo apt-get clean
```

After cleaning the cache, use one of the following upgrade commands to upgrade CDH.

Precise:

```
$ sudo apt-get upgrade -t precise-cdh4
```

Lucid:

```
$ sudo apt-get upgrade -t lucid-cdh4
```

Squeeze:

```
$ sudo apt-get upgrade -t squeeze-cdh4
```

- Use a custom repository. This process can be more complicated, but enables updating CDH components for hosts that are not connected to the Internet. You can create your own repository, as described in [Understanding Custom Installation Solutions](#). Creating your own repository is necessary if you are upgrading a cluster that does not have access to the Internet.

If you used a custom repository to complete the installation of your current files and now you want to update using a custom repository, the details of the steps to complete the process are variable. In general, begin by updating any existing custom repository that you will use with the installation files you want to use. This can be completed in a variety of ways. For example, you might use `wget` to copy the necessary installation files. Once the installation files have been updated, use the custom repository you established for the initial installation to update CDH.

OS	Command
RHEL	<p>Ensure you have a custom repo that is configured to use your internal repository. For example, if you could have custom repo file in <code>/etc/yum.conf.d/</code> called <code>cdh_custom.repo</code> in which you specified a local repository. In such a case, you might use the following commands:</p> <pre>\$ sudo yum clean all \$ sudo yum update 'cloudera-*</pre>
SLES	<p>Use commands such as the following to clean cached repository information and then update only the CDH components:</p> <pre>\$ sudo zypper clean --all \$ sudo zypper up -r http://internalserver.example.com/path_to_cdh_repo</pre>
Ubuntu or Debian	<p>Use a command that targets upgrade of your CDH distribution using the custom repository specified in your apt configuration files. These files are typically either the <code>/etc/apt/apt.conf</code> file or in various files in the <code>/etc/apt/apt.conf.d/</code> directory. Information about your custom repository must be included in the repo files. The general form of entries in Debian/Ubuntu is:</p> <pre>deb http://server.example.com/directory/ dist-name pool</pre> <p>For example, the entry for the default repo is:</p> <pre>deb http://us.archive.ubuntu.com/ubuntu/ precise universe</pre> <p>On a Debian/Ubuntu system, use commands such as the following to clean cached repository information and then update only the CDH components:</p> <pre>\$ sudo apt-get clean \$ sudo apt-get upgrade -t your_cdh_repo</pre>

Upgrade the Hive Metastore Database

Required if you are upgrading from an earlier version of CDH 4 to CDH 4.2 or higher.

1. Go to the Hive service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
4. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade the Oozie ShareLib

1. Go to the Oozie service.
2. Select **Actions > Start** and click **Start** to confirm.
3. Select **Actions > Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Restart the Services

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Restart**.

2. Click the **Restart** button in the confirmation pop-up that appears. The **Command Details** window shows the progress of starting services.

Configure Cluster CDH Version for Package Installs

If you have installed CDH as a package, after an installation or upgrade, make sure that the cluster CDH version matches the package CDH version, using the procedure in [Configuring the CDH Version of a Cluster](#) on page 113. If the cluster CDH version does not match the package CDH version, Cloudera Manager incorrectly enables and disables service features based on the cluster's configured CDH version.

Deploy Client Configuration Files

1. On the **Home > Status** tab, click



to the right of the cluster name and select **Deploy Client Configuration**.

2. Click the **Deploy Client Configuration** button in the confirmation pop-up that appears.

Configuring the CDH Version of a Cluster

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

Cloudera Manager has version-specific features based on minor and maintenance versions. For example, the Sqoop 2 service and the HiveServer2 and WebHCat roles are only available for CDH 4.2.0 and higher. These behaviors are controlled by what is configured in Cloudera Manager, not what is actually installed on the hosts. The versions should be in sync, and is the case if parcels are used.

In package-based clusters, you can manually upgrade CDH packages. For example, you can upgrade the packages from CDH 4.1.0 to CDH 4.2.1. However, in previous releases, Cloudera Manager did not detect this change and behaved as if the cluster was 4.1.0. In such cases, it would not display Sqoop 2 as a service. You would have to set the CDH version manually using the cluster Configure CDH Version action.

Cloudera Manager now sets the CDH version correctly. However, if you have an older Cloudera Manager and forgot to set the version, and then upgraded to latest Cloudera Manager, you need to set the version.

To inform Cloudera Manager of the CDH version, select **ClusterName > Configure CDH Version**. In the dialog box, Cloudera Manager displays the installed CDH version, and asks for confirmation to configure itself with the new version. The dialog box will also detect if a major upgrade was done, and direct you to use the major upgrade flow documented in [Upgrading from CDH 4 Packages to CDH 5 Packages](#) on page 95.

Performing a Rolling Upgrade on a CDH 4 Cluster

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



Important: This feature is available only with a Cloudera Enterprise license. It is not available in Cloudera Express. For information on Cloudera Enterprise licenses, see [Managing Licenses](#).

The rolling upgrade feature takes advantage of parcels and the HDFS high availability to enable you to upgrade your cluster software and restart the upgraded services without taking the entire cluster down. You must have HDFS high availability enabled to perform a rolling upgrade.

This page described how to perform a rolling upgrade between minor versions of CDH 4. For rolling upgrade between CDH 5 versions, see [Upgrading to CDH 5.x Using a Rolling Upgrade](#) on page 55.

Upgrading CDH and Managed Services Using Cloudera Manager

You cannot perform a rolling upgrade from CDH 4 to CDH 5 because of incompatibilities between the two major versions. Instead, follow the instructions for a full upgrade at [Upgrading from CDH 4 to CDH 5 Parcels](#) on page 87.

A rolling upgrade involves two steps:

1. Download, distribute, and activate the parcel for the new software you want to install.
2. Perform a [rolling restart](#) to restart the services in your cluster. You can do a rolling restart of individual services, or if you have high availability enabled, you can perform a restart of the entire cluster. Cloudera Manager will manually fail over your NameNode at the appropriate point in the process so that your cluster will not be without a functional NameNode.

To avoid a high number of alerts during the upgrade process, you can enable [maintenance mode](#) on your cluster before you start the upgrade. This stops email alerts and SNMP traps from being sent, but does not stop checks and configuration validations. Exit maintenance mode when you have finished the upgrade to reenable Cloudera Manager alerts.



The steps to perform a rolling upgrade of a cluster are as follows:

Ensure High Availability is Enabled

To enable high availability, see [HDFS High Availability](#) for instructions. You do not need to enable automatic failover for rolling restart to work, though you can enable it if you want. Automatic failover does not affect the rolling restart operation. If you have JobTracker high availability configured, Cloudera Manager will fail over the JobTracker during the rolling restart, but this is not a requirement for performing a rolling upgrade.

Download, Distribute, and Activate Parcels

1.

In the Cloudera Manager Admin Console, click the Parcels indicator in the top navigation bar ( or ) to go to the Parcels page.

2. In the parcels page, click **Download** for the version(s) you want to download. If the parcel you want is not shown here — for example, you want to upgrade to version of CDH that is not the most current version — you can make additional parcel repos available through the [parcel settings](#) page. If your Cloudera Manager server does not have Internet access, you can obtain the required parcel file(s) and put them into the local repository. See [Creating and Using a Parcel Repository for Cloudera Manager](#) for more details.
3. When the download has completed, click **Distribute** for the version you downloaded.
4. When the parcel has been distributed and unpacked, the button will change to say **Activate**.
5. Click **Activate**. You are asked if you want to restart the cluster. *Do not restart the cluster at this time.*
6. Click **Close**.

Upgrade the Hive Metastore Database

Required if you are upgrading from an earlier version of CDH 4 to CDH 4.2 or higher.

1. Go to the Hive service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Select **Actions > Upgrade Hive Metastore Database Schema** and click **Upgrade Hive Metastore Database Schema** to confirm.
4. If you have multiple instances of Hive, perform the upgrade on each metastore database.

Upgrade the Oozie ShareLib

1. Go to the Oozie service.
2. Select **Actions > Start** and click **Start** to confirm.
3. Select **Actions > Install Oozie ShareLib** and click **Install Oozie ShareLib** to confirm.

Upgrade Sqoop

1. Go to the Sqoop service.

2. Select **Actions** > **Stop** and click **Stop** to confirm.
3. Select **Actions** > **Upgrade Sqoop** and click **Upgrade Sqoop** to confirm.

Upgrade the Sentry Database

Required for the following upgrades:

- CDH 5.1 to 5.2 or higher
 - CDH 5.2 to 5.3 or higher
 - CDH 5.4 to 5.5 or higher
1. Go to the Sentry service.
 2. Select **Actions** > **Stop** and click **Stop** to confirm.
 3. Select **Actions** > **Upgrade Sentry Database Tables** and click **Upgrade Sentry Database Tables** to confirm.

Upgrade Spark

1. Go to the Spark service.
2. Select **Actions** > **Stop** and click **Stop** to confirm.
3. Select **Actions** > **Install Spark JAR** and click **Install Spark JAR** to confirm.

Restart the Cluster

1. On the **Home** > **Status** tab, click



to the right of the cluster name and click **Rolling Restart** to proceed with a rolling restart. Rolling restart is available only if high availability is enabled. Click **Restart** to perform a normal restart. Services that do not support rolling restart will undergo a normal restart, and will not be available during the restart process.

2. For a rolling restart, a pop-up allows you to choose which services you want to restart, and presents caveats to be aware of for those services that can undergo a rolling restart.



Note: If you have just upgraded your Cloudera Manager deployment to 4.6, and are now doing a rolling upgrade of your cluster, you must ensure that MapReduce is restarted *before* the rest of your services, or the restart may fail. This is necessary to ensure the MapReduce configuration changes are propagated.

Further, if you are upgrading from CDH 4.1 with Impala to CDH 4.2 or 4.3, you must restart MapReduce before Impala restarts (by default Impala is restarted before MapReduce).

The workaround is to perform a restart of MapReduce alone as the first step, then perform a cluster restart of the remaining services.

3. Click **Confirm** to start the rolling restart.

Remove CDH 4 Packages

If your previous installation of CDH was done using packages, remove those packages on all hosts on which you installed the parcels and refresh the symlinks so that clients will run the new software versions. *Skip this step if your previous installation was using parcels.*

1. If Hue is configured to use SQLite as its database:
 - a. Stop the Hue service.
 - b. Back up the `desktop.db` to a temporary location before deleting the old Hue Common package. The location of the database can be found in the Hue service **Configuration** tab under **Service** > **Database** > **Hue's Database Directory**.



Important: Removing the Hue Common package will remove your Hue database; if you do not back it up you may lose all your Hue user account information.

2. Uninstall the CDH packages on each host:

- **Not including Impala and Search**

Operating System	Command
RHEL	<code>\$ sudo yum remove hadoop hue-common bigtop-jsvc bigtop-tomcat</code>
SLES	<code>\$ sudo zypper remove hadoop hue-common bigtop-jsvc bigtop-tomcat</code>
Ubuntu or Debian	<code>\$ sudo apt-get purge hadoop hue-common bigtop-jsvc bigtop-tomcat</code>

- **Including Impala and Search**

Operating System	Command
RHEL	<code>\$ sudo yum remove hadoop hue-common impala-shell solr-server 'bigtop-*</code>
SLES	<code>\$ sudo zypper remove hadoop hue-common impala-shell solr-server 'bigtop-*</code>
Ubuntu or Debian	<code>\$ sudo apt-get purge hadoop hue-common impala-shell solr-server 'bigtop-*</code>

3. Restart all the Cloudera Manager Agents to force an update of the symlinks to point to the newly installed components on each host:

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

Restore Hue Database

If you [removed CDH 4 packages](#), restore the Hue database back up.

1. Go to the Hue service.
2. Select **Actions > Stop** and click **Stop** to confirm.
3. Copy the backup from the temporary location to the newly created Hue database directory:
`/opt/cloudera/parcels/CDH-4.x.0-x.cdh4.x.0.p0.xx/share/hue/desktop.`
4. Restart the Hue service.

Upgrading CDH 3



Warning:

- Cloudera Manager 4 and CDH 4 have reached End of Maintenance (EOM) on August 9, 2015. Cloudera does not support or provide updates for Cloudera Manager 4 and CDH 4 releases.
- Cloudera Manager 3 and CDH 3 have reached End of Maintenance (EOM) on June 20, 2013. Cloudera does not support or provide updates for Cloudera Manager 3 and CDH 3 releases.

To upgrade CDH 3 to CDH 4 with Cloudera Manager 4, follow the instructions at [Upgrading CDH 3 to CDH 4 in a Cloudera Manager Deployment](#).

Rolling Back a CDH 4-to-CDH 5 Upgrade

You can roll back an upgrade from CDH 4 to CDH 5. The rollback restores your CDH cluster to the state it was in before the upgrade, including Kerberos and TLS/SSL configurations. Any data created after the upgrade is lost.

In a typical upgrade, you first upgrade Cloudera Manager from version 4.x to version 5.x, and then you use the upgraded version of Cloudera Manager 5 to upgrade CDH 4 to CDH 5. (See [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54.) If you want to roll back this upgrade, follow these steps to roll back your cluster to its state prior to the upgrade.



Important: Follow all of the steps in the order presented in this topic. Cloudera recommends that you read through the backup and rollback steps before starting the backup process. You may want to create a detailed plan to help you anticipate potential problems.

High-Level Steps

1. Before the upgrade, [back up specified directories, files, and databases used in your CDH cluster](#).
2. To roll back your cluster to its state before the upgrade, perform the [rollback steps on the hosts in your CDH cluster](#).
3. Use Cloudera Manager 5 to start and manage the CDH 4 cluster. You can use Cloudera Manager 5 to manage both CDH 4 and CDH 5 clusters. (If required, you can also [roll back Cloudera Manager 5 to Cloudera Manager 4](#).)

Limitations

The rollback procedure has the following limitations:

- If you have [finalized the HDFS upgrade](#), you cannot roll back your cluster.
- Configuration changes, including the addition of new services or roles after the upgrade, are not retained after rolling back Cloudera Manager.
- Cloudera recommends that you not make configuration changes or add new services and roles until you have finalized the HDFS upgrade and no longer require the option to roll back your upgrade.
- If your cluster is configured to use HBase replication, data written to HBase after the upgrade might not be replicated to peers when you start your rollback. This topic does not describe how to determine which, if any, peers have the replicated data and how to roll back that data. For more information about HBase replication, see [HBase Replication](#).



Note:

- Hadoop version 2.0, which is included with CDH 4, does not support HDFS rollback for clusters on which high availability is enabled for HDFS. If your CDH 4 cluster has high availability enabled for HDFS, you can temporarily reconfigure your cluster without high availability before proceeding with the rollback. After the rollback, you can re-enable high availability for HDFS. Procedures for these reconfigurations are provided in [the rollback procedures](#).
- Because of an HDFS bug, rollback fails on DataNodes because the DataNode `storageID` format changed between the Hadoop versions used in CDH 4 and CDH 5. The workaround is described in [the rollback procedures](#).

Backing Up Before Upgrading from CDH 4 to CDH 5

Backing up CDH components before upgrading your Cloudera Manager and CDH software provides a way to roll back the upgrade. This topic provides procedures to back up your cluster so that you can roll back your cluster to its pre-upgrade state.

Backup Steps

1. [Preparing to Back Up](#) on page 118
2. [Stopping the Cluster](#) on page 118
3. [Backing Up CDH 4 and Cloudera Manager Repository Files](#) on page 118
4. [Backing Up ZooKeeper](#) on page 119
5. [Backing Up HDFS \(With High Availability\)](#) on page 119
6. [Backing Up HDFS \(Without High Availability\)](#) on page 119
7. [Backing Up HBase](#) on page 119
8. [Backing Up Hive](#) on page 119
9. [Backing Up Oozie](#) on page 119
10. [Backing Up Search](#) on page 120
11. [Backing Up Sqoop 2](#) on page 120
12. [Backing Up Hue](#) on page 120
13. [Backing Up Cloudera Manager](#) on page 120
14. [Backing Up Other CDH Components](#) on page 120

Preparing to Back Up

Because many storage locations are configurable, you may need to use the Cloudera Manager Admin Console to determine the location of files you need to back up. Where applicable, the parameter names that specify these locations are provided in the backup sections for each component. If you have not changed these values for your cluster, use the provided default values. To find these parameter values:

1. Open the Cloudera Manager Admin console.
2. Go to the service where you need to look up a parameter (for example, HDFS, HBase, or ZooKeeper).
3. Click the **Configuration** tab.
4. Enter the name of the parameter in the search box.

The parameter and its value display on the right.

For some services, you back up data stored in relational databases such as Oracle, MariaDB, MySQL, or PostgreSQL. [See the documentation](#) for those products to learn how to back up and restore the databases.



Important:

- As you back up the required files described in this topic, record which hosts the backups come from (or back up the files to the same host).
- After starting your backups, do not add additional components or change any configurations until your upgrade is successfully completed or rolled back.
- Complete all backup steps before starting the upgrade to CDH 5.

Stopping the Cluster

Stop the CDH cluster before performing the backup:

1. Go to the **Home** page.
2. In the drop-down list next to your cluster, select **Stop**.

Backing Up CDH 4 and Cloudera Manager Repository Files

If your cluster was installed using *packages*, back up the CDH 4 and Cloudera Manager repository files on all hosts from the system repository directory. If your cluster was installed using Cloudera *parcels*, back up only the Cloudera Manager repository file. Following are the typical locations for the repository directories:

Operating System	Path
RHEL	/etc/yum.repos.d

Operating System	Path
SLES	<code>/etc/zypp/repos.d</code>
Ubuntu or Debian	<code>/etc/apt/sources.list.d</code>

For example, on a RHEL or similar system, back up the files in `/etc/yum.repos.d` that have `cloudera` as part of their name.

Backing Up ZooKeeper

On all ZooKeeper hosts, back up the ZooKeeper data directory specified with the `dataDir` property in the ZooKeeper configuration. The default location is `/var/lib/zookeeper`.

Record the permissions of the files and directories; you will need these to roll back ZooKeeper.

Backing Up HDFS (With High Availability)

Follow this procedure to back up an HDFS deployment that has been configured for high availability.

1. On both NameNode hosts, back up one of the NameNode data directories specified with the `dfs.namenode.name.dir` property.
2. On each JournalNode, back up the JournalNode edits directory specified by the `dfs.journalnode.edits.dir` property. Note which JournalNode host the backup comes from.
3. Back up the `VERSION` files for each DataNode, noting which DataNode you are backing up. There may be multiple data directories in each node, but you need to back up only one of them on each DataNode. The location of the data directories is specified with the `dfs.datanode.data.dir` property. The `VERSION` file is located in the `current` subdirectory. You will use the version files to get the `storageID` when you perform the rollback steps; for example (using the default path): `/data/dfs/dn/current/VERSION`. You only need this `storageID` when rolling back the DataNodes; copying the `VERSION` file is suggested as a convenience.

Backing Up HDFS (Without High Availability)

Use this procedure to back up an HDFS deployment that has not been configured for high availability.

1. On the NameNode host, back up one of the NameNode data directories specified with the `dfs.namenode.name.dir` property.
2. Back up the `VERSION` files for each DataNode, noting which DataNode you are backing up. There may be multiple data directories in each node, but you need to back up only one of them on each DataNode. The location of the data directories is specified with the `dfs.datanode.data.dir` property. The `VERSION` file is located in the `current` subdirectory. You will use the version files to get the `storageID` when you perform the rollback steps; for example (using the default path): `/data/dfs/dn/current/VERSION`. You only need this `storageID` when rolling back the DataNodes; copying the `VERSION` file is suggested as a convenience.

Backing Up HBase

Because the rollback procedure also rolls back HDFS, the data in HBase is also rolled back. In addition, HBase metadata stored in ZooKeeper is recovered as part of the ZooKeeper rollback procedure.

If your cluster is configured to use HBase replication, Cloudera recommends that you document all replication peers. If necessary (for example, because the HBase `znode` has been deleted), you can roll back HBase as part of the HDFS rollback without the ZooKeeper metadata. This metadata can be reconstructed in a fresh ZooKeeper installation, with the exception of the replication peers, which you must add back. For information on enabling HBase replication, listing peers, and adding a peer, see [HBase Replication](#) in the CDH 4 documentation.

Backing Up Hive

Back up the database that backs the Hive metastore. See [Backing up Databases](#) on page 121.

Backing Up Oozie

Back up the Oozie database. See [Backing up Databases](#) on page 121.

Backing Up Search

On each Solr node, back up the contents of the Solr Data directory and record the permissions for the directory. This location is specified with the **Solr Data Directory** property. The default location is:

```
/var/lib/solr
```

Search data on ZooKeeper is restored as part of the ZooKeeper rollback.

Backing Up Sqoop 2

If you are not using the default embedded Derby database for Sqoop 2, [back up the database](#) you have configured for Sqoop 2. Otherwise, back up the `repository` subdirectory of the Sqoop 2 metastore directory. This location is specified with the **Sqoop 2 Server Metastore Directory** property. The default location is: `/var/lib/sqoop2`. For this default location, Derby database files are located in `/var/lib/sqoop2/repository`.

Backing Up Hue

1. Back up the Hue database. See [Backing up Databases](#) on page 121.
2. Back up the app registry file, `<HUE_HOME>/app.reg`, where `HUE_HOME` is the location of your Hue installation. For package installs, this is usually `/usr/lib/hue`; for parcel installs, this is usually, `/opt/cloudera/parcels/<parcel version>/lib/hue/`.

Backing Up Cloudera Manager

1. Stop Cloudera Management Services using Cloudera Manager:
 - a. Select **Clusters > Cloudera Management Service**.
 - b. Select **Actions > Stop**.
2. Stop Cloudera Manager Server by running the following command on the Cloudera Manager Server host:

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

3. On the host where Cloudera Manager Server is running, back up the `/etc/cloudera-scm-server/db.properties` file.
4. On the host where the Event Server role is configured to run, back up the contents of the directory specified with the **Event Server Index Directory** property (the default value is `/var/lib/cloudera-scm-eventserver`).
5. Back up the `/etc/cloudera-scm-agent/config.ini` file on each host in the cluster.
6. Back up the following Cloudera Manager-related databases; see [Backing up Databases](#) on page 121:
 - Cloudera Manager Server
 - Activity Monitor (depending on your deployment, this role may not be installed)
 - Reports Manager
 - Service Monitor
 - Host Monitor
 - Navigator Audit Server
 - Navigator Metadata Server

Backing Up Other CDH Components

No backups are required for the following components:

- MapReduce
- YARN
- Spark
- Pig
- Sqoop
- Impala

Backing up Databases

Several steps in the backup procedures require you to back up various databases used in a CDH cluster. The steps for backing up and restoring databases differ depending on the database vendor and version you select for your cluster and are beyond the scope of this document.

See the following vendor resources for more information:

- **MariaDB 5.5:** <http://mariadb.com/kb/en/mariadb/backup-and-restore-overview/>
- **MySQL 5.5:** <http://dev.mysql.com/doc/refman/5.5/en/backup-and-recovery.html>
- **MySQL 5.6:** <http://dev.mysql.com/doc/refman/5.6/en/backup-and-recovery.html>
- **PostgreSQL 8.4:** <https://www.postgresql.org/docs/8.4/static/backup.html>
- **PostgreSQL 9.2:** <https://www.postgresql.org/docs/9.2/static/backup.html>
- **PostgreSQL 9.3:** <https://www.postgresql.org/docs/9.3/static/backup.html>
- **Oracle 11gR2:** http://docs.oracle.com/cd/E11882_01/backup.112/e10642/toc.htm

Procedure for Rolling Back a CDH 4-to-CDH 5 Upgrade

You can roll back to CDH 4 after upgrading to CDH 5 only if the [HDFS upgrade has not been finalized](#). The rollback restores your CDH cluster to the state it was in before the upgrade, including Kerberos and TLS/SSL configurations. Data created after the upgrade is lost.



Important: When performing these rollback steps, use backups taken before you started the upgrade. For steps where you need to restore the contents of a directory, clear the contents of the directory before copying the backed-up files to the directory. If you fail to do this, artifacts from the original upgrade can cause problems if you attempt the upgrade again after the rollback.

Rollback Steps

1. [Determining the Last Active NameNode](#) on page 121.
2. [Downgrading the Software](#) on page 121.
3. [Rolling Back ZooKeeper](#) on page 123.
4. [Rolling Back HDFS](#) on page 123.
5. [Re-enabling HDFS with High Availability](#) on page 125 (optional).
6. [Rolling Back HBase](#) on page 125.
7. [Rolling Back Hive](#) on page 126.
8. [Rolling Back Oozie](#) on page 126.
9. [Rolling Back Search](#) on page 126.
10. [Rolling Back Sqoop 2](#) on page 126.
11. [Rolling Back Hue](#) on page 126.
12. [Rolling Back Cloudera Manager](#) on page 126.

Determining the Last Active NameNode

If your cluster has high availability for HDFS enabled, determine which NameNode host was last active:

1. In Cloudera Manager, select **Clusters > HDFS > Instances**.

A list of role types and hosts displays.

2. Look for the Role Type **NameNode (Active)**.

The NameNode that is not active is called the *standby* NameNode and displays in the list of Role Types as **NameNode (Standby)**.

Downgrading the Software

1. Stop the cluster:

- a. On the **Home > Status** tab, click



to the right of the cluster name and select **Stop**.

- b. Click **Stop** in the confirmation screen. The **Command Details** window shows the progress of stopping services.

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

2. Downgrade the JDK, if necessary.

When you upgraded your cluster to CDH 5, you were required to also upgrade the JDK. If you are rolling back your cluster to Cloudera Manager version 4.6.x or lower and CDH version 4.3 .x or lower, Cloudera recommends that you downgrade your JDK on all hosts to the version deployed before the upgrade, or to JDK 1.6. You can also choose to run your cluster using the version of the JDK you installed during the upgrade. To verify that the JDK you choose is supported by Cloudera Manager and CDH, see [CDH and Cloudera Manager Supported JDK Versions](#).

You can have two versions of a JDK on a host machine, but you must make sure the correct version is used by the software. See: [Java Development Kit Installation](#).

3. Depending on whether your cluster was installed using *parcels* or *packages*, do one of the following:

• Parcels

1. Log in to the Cloudera Manager Admin Console.

2. Select **Hosts > Parcels**.

A list of parcels displays.

3. Locate the CDH 4 parcel and click **Activate**. (This automatically deactivates the CDH 5 parcel.) See [Activating a Parcel](#) for more information. If the parcel is not available, use the **Download** button to download the parcel.

4. If you include any additional components in your cluster, such as Search or Impala, click **Activate** for those parcels.



Important:

Do not start any services.

If you accidentally restart services, stop your cluster before proceeding.

• Packages

1. Log in as a privileged user to all hosts in your cluster.

2. Run the following command to uninstall CDH 5:

Operating System	Command
RHEL	\$ sudo yum remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc
SLES	\$ sudo zypper remove bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc
Ubuntu or Debian	\$ sudo apt-get purge bigtop-jsvc bigtop-utils bigtop-tomcat hue-common sqoop2-client hbase-solr-doc solr-doc

3. Remove the CDH 5 repository files from the system repository directory. For example, on a RHEL or similar system, remove all files in `/etc/yum.repos.d` that have `cloudera` as part of the name. (Make sure that you have backed up these files, as instructed in [Backing Up CDH 4 and Cloudera Manager Repository Files](#) on page 118.)

4. Restore the CDH 4 repository files that you previously backed up to the repository directory.
5. Re-install the CDH 4 packages using the same installation path you used for the initial installation; see [Installing Cloudera Manager and CDH](#). Repeat only the installation steps and not the configuration steps. (The configurations are already stored in Cloudera Manager.) Make sure that you include any additional components used in your cluster, such as MapReduce 1, YARN, Spark, Pig, Sqoop, or Impala.

Rolling Back ZooKeeper

1. Restore the contents of the `dataDir` on each ZooKeeper server. These files are located in a directory specified with the `dataDir` property in the ZooKeeper configuration. The default location is `/var/lib/zookeeper`.
2. Make sure that the permissions of all the directories and files are as they were before the upgrade.
3. Start ZooKeeper using Cloudera Manager.

Rolling Back HDFS

You cannot roll back HDFS while high availability is enabled. The rollback procedure in this topic creates a temporary configuration without high availability. Regardless of whether high availability is enabled, follow the steps in this section to create the temporary configuration, and to re-enable high availability, follow the steps in [Re-enabling HDFS with High Availability](#) on page 125.

1. Create a rollback configuration directory on each host that has the HDFS role; for example `/etc/hadoop/conf.rollback`.
2. Create a `core-site.xml` file in this directory on all hosts with the HDFS role. The `<value>` element in this file references the NameNode host. For HDFS with high availability enabled, choose the last active NameNode host (see [Determining the Last Active NameNode](#) on page 121). The file needs to contain only the following:

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://NameNode_host:port</value>
    <!-- For example:
      hdfs://a1234.cloudera.com:8020
    -->
  </property>
</configuration>
```

3. Create an `hdfs-site.xml` file in the rollback configuration directory on all hosts with the HDFS role. The file specifies values for the following properties:

- `dfs.namenode.name.dir`
- `dfs.datanode.data.dir`
- `dfs.namenode.checkpoint.dir`

You can omit the `dfs.namenode.checkpoint.dir` property if high availability was enabled for HDFS in your cluster.

To find values you need to create the `hdfs-site.xml` file:

1. Open Cloudera Manager.
2. Go to the HDFS service.
3. Click the **Configuration** tab.
4. Enter the property name in the search field.
5. Copy each path defined for the property into the `<value>` element for the property. Precede each path with `file://` and separate each directory path with a comma.

For example:

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///data/1/dfs/nn,file:///data/2/dfs/nn</value>
```

```
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///data/1/dfs/dn,file:///data/2/dfs/dn,file:///data/3/dfs/dn,
    file:///data/4/dfs/dn,file:///data/5/dfs/dn
  </value>
</property>
<property>
  <name>dfs.namenode.checkpoint.dir</name>
  <value>file:///data/1/dfs/snn</value>
</property>
</configuration>
```

4. Restore the `storageID` values from the `VERSION` files. On each `DataNode`, edit the `VERSION` file in each data directory and replace the value of the `storageID` field with the value from the `VERSION` file you backed up from the same `DataNode`. The `storageID` is the same for each data directory on a `DataNode` host.



Important:

Restore the `storageID` on *all data directories* on *all DataNodes*.

Do not restore the entire `VERSION` file.

5. Copy the `/etc/hadoop/conf/log4j.properties` file to the `/etc/hadoop/conf.rollback` directory (the rollback configuration directory you created previously).
6. Verify that the cluster is now running CDH 4 by running the following command on all Hadoop nodes in your cluster:

```
hadoop version
```

The version number appears right after the string `cdh`. For example:

```
$ hadoop version
Hadoop 2.0.0-cdh4.7.1
...
```

7. Run the following command on the `NameNode` host. (If high availability is enabled for HDFS, run the command on the active `NameNode`.)

```
sudo -u hdfs hdfs --config /etc/hadoop/conf.rollback namenode -rollback
```

This command also starts the `NameNode`.

8. While the `NameNode` is running, run the following command on all `DataNode` hosts:

```
sudo -u hdfs hdfs --config /etc/hadoop/conf.rollback datanode -rollback
```

This command also starts the `DataNodes`.

9. If your cluster uses a `Secondary NameNode`, do the following on the `Secondary NameNode` host:

- a. Delete the directory defined with the `dfs.namenode.checkpoint.dir` parameter.
- b. Run the following command while the active `NameNode` and all `DataNodes` are running:

```
sudo -u hdfs hdfs --config /etc/hadoop/conf.rollback secondarynamenode -format
```

10. When rollback is completed, stop all daemons by typing `CTRL-C` in all open terminal sessions (the `NameNode` hosts and `DataNode` hosts on which you ran the `hdfs` commands). You can monitor the progress of the rollback by opening the `NameNode` web interface in a web browser at `NameNode_Host:50070`. On that page, when all the blocks are reported and the `NameNode` no longer reports that it is in **Safemode**, the rollback is complete.

11. If your cluster does not have high availability enabled for HDFS, use the Cloudera Manager Admin Console to start the HDFS service and then continue your cluster rollback with [Rolling Back HBase](#) on page 125:

- a. Go to the HDFS service.
- b. Select **Actions > Start**.

If your cluster has high availability enabled for HDFS, continue with the next section to remove the temporary non-HA configuration and re-enable high availability.

Re-enabling HDFS with High Availability

The rollback steps performed so far have restored all of the DataNodes and one of the NameNodes. In a high-availability configuration, you also need to roll back the other (standby) NameNode and the JournalNodes.

1. On each JournalNode, restore the backed-up edits directory to the same location.
2. On the standby NameNode (the NameNode host that has not been rolled back), delete the directories specified in the `dfs.namenode.name.dir` property. Failing to delete these directories causes the rollback to fail.
3. Using Cloudera Manager, start the last active NameNode:
 - a. Go to the HDFS service.
 - b. Click the **Instances** tab.
 - c. Select **NameNode(Active)** from the list of Role Types.
 - d. Select **Actions for Selected > Start**.
4. Using Cloudera Manager, bootstrap the standby NameNode:
 - a. Go to the HDFS service.
 - b. Click the **Instances** tab.
 - c. Select **NameNode(Standby)**.
 - d. Select **Actions > Bootstrap Standby NameNode**.

The `fsImage` is restored from the last active NameNode to the standby NameNode.

5. Restart the HDFS service:
 - a. Go to the HDFS service.
 - b. Select **Actions > Restart**.

Rolling Back HBase

No additional steps are required to roll back HBase. When you rolled back HDFS, all of the HBase data was also rolled back.

Using Cloudera Manager, start HBase:

1. Go to the HBase service.
2. Select **Actions > Start**.

If you encounter errors when starting HBase, delete the `znode` in ZooKeeper and then start HBase again:

1. In Cloudera Manager, look up the value of the `zookeeper.znode.parent` property. The default value is `/hbase`.
2. Connect to the ZooKeeper ensemble by running the following command from any HBase gateway host:

```
zookeeper-client -server zookeeper_ensemble
```

To find the value to use for `zookeeper_ensemble`, open the `/etc/hbase/conf/hbase-site.xml` file on any HBase gateway host. Use the value of the `hbase.zookeeper.quorum` property.

**Note:**

If you have deployed a secure cluster, you must connect to ZooKeeper using a client `jaas.conf` file. You can find such a file in an HBase process directory (`/var/run/cloudera-scm-agent/process/`). Specify the `jaas.conf` using the JVM flags by running the following commands in the ZooKeeper client:

```
CLIENT_JVMFLAGS=  
"-Djava.security.auth.login.config=/var/run/cloudera-scm-agent/process/HBase_process_directory/jaas.conf"  
zookeeper-client -server <zookeeper_ensemble>
```

The ZooKeeper command-line interface opens.

3. Enter the following command:

```
rmr /hbase
```

Rolling Back Hive

Restore the Hive metastore database from your backup. [See the documentation for your database](#) for details.

Rolling Back Oozie

Restore the Oozie database from your backup. [See the documentation for your database](#) for details.

Rolling Back Search

Restore the contents of the `/var/lib/solr` directory from your backup. Ensure that the file permissions are the same as they were before rolling back. (The permissions are typically `solr:solr`.)

Rolling Back Sqoop 2

Restore the Sqoop 2 database from your backup. [See the documentation for your database](#) for details.

If you are not using the default embedded Derby database for Sqoop 2, [restore the database](#) you have configured for Sqoop 2. Otherwise, restore the `repository` subdirectory of the Sqoop 2 metastore directory from your backup. This location is specified with the **Sqoop 2 Server Metastore Directory** property. The default location is `/var/lib/sqoop2`. For this default location, Derby database files are located in `/var/lib/sqoop2/repository`.

Rolling Back Hue

1. Restore the Hue database from your backup.
2. Restore the file, `app.reg`, from your backup. Place it in the location of your Hue installation (referred to as `HUE_HOME`). For package installs, this is usually `/usr/lib/hue`; for parcel installs, this is usually, `/opt/cloudera/parcels/<parcel version>/lib/hue/`.
3. Using Cloudera Manager, install the Beeswax role:
 - a. Go to the Hue service.
 - b. Select the **Instances** tab.
 - c. Click **Add**.
 - d. Locate the row for the host with the Hue Server role and select **Beeswax Server**.
 - e. Click **Continue**.

Rolling Back Cloudera Manager

After you complete the rollback steps, your cluster is using Cloudera Manager 5 to manage your CDH 4 cluster. You can continue to use Cloudera Manager 5 to manage your CDH 4 cluster, or you can downgrade to Cloudera Manager 4 by following these steps:

1. Stop your CDH cluster using Cloudera Manager:

- a. Go to the **Home** page.
- b. In the drop-down list next to your cluster, select **Stop**.

2. Stop Cloudera Management Services:

- a. Select **Clusters > Cloudera Management Service**.
- b. Select **Actions > Stop**

3. Stop Cloudera Manager Server by running the following command on the Cloudera Manager Server host:

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

4. Stop the Cloudera Manager Agents by running the following command on all hosts in your cluster:

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

5. Downgrade the JDK, if necessary.

If you are rolling back to Cloudera Manager version 4.6.x or lower, downgrade the version of your JDK to the version deployed before the upgrade, or the latest Oracle upgrade for JDK 1.6. You can also choose to run Cloudera Manager using the version of the JDK you installed during the upgrade. To verify that the JDK you choose is supported by Cloudera Manager and CDH, see [CDH and Cloudera Manager Supported JDK Versions](#).

You can keep two versions of a JDK on a host machine, but you must make sure the correct version is used by the software. See [Java Development Kit Installation](#) and [Configuring a Custom Java Home Location](#) for information about specifying a JDK.

6. Downgrade the software:

- a. Run the following command on the Cloudera Manager Server host:

Operating System	Command
RHEL	\$ sudo yum remove cloudera-manager-server
SLES	\$ sudo zypper remove cloudera-manager-server
Ubuntu or Debian	\$ sudo apt-get purge cloudera-manager-server

- b. Run the following command on all hosts:

Operating System	Command
RHEL	\$ sudo yum remove cloudera-manager-daemons cloudera-manager-agent
SLES	\$ sudo zypper remove cloudera-manager-daemons cloudera-manager-agent
Ubuntu or Debian	\$ sudo apt-get purge cloudera-manager-daemons cloudera-manager-agent

- c. Restore the Cloudera Manager repository files that you previously backed up to the Cloudera repository directory.
- d. Run the following commands on all hosts:

Operating System	Command
RHEL	<pre>\$ sudo yum clean all \$ sudo yum install cloudera-manager-agent</pre>
SLES	<pre>\$ sudo zypper refresh -s \$ sudo zypper install cloudera-manager-agent</pre>
Ubuntu or Debian	<pre>\$ sudo apt-get update \$ sudo apt-get install cloudera-manager-agent</pre>

e. Run the following commands on the Cloudera Manager server host:

Operating System	Command
RHEL	<pre>\$ sudo yum install cloudera-manager-server</pre>
SLES	<pre>\$ sudo zypper install cloudera-manager-server</pre>
Ubuntu or Debian	<pre>\$ sudo apt-get install cloudera-manager-server</pre>

7. Restore the following Cloudera Manager databases:

- Cloudera Manager Server
- Activity Monitor (depending on your deployment, this role may not be installed)
- Reports Manager
- Service Monitor
- Host Monitor
- Navigator Audit Server
- Navigator Metadata Server

[See the documentation for your database](#) for details.



Important: Restore the databases to their pre-upgrade state. If the Cloudera Manager databases are restored in a way that leaves tables that were created during the upgrade, there could be problems if you attempt to upgrade Cloudera Manager again after the rollback.

8. On the Cloudera Manager Server, restore the following from your backup to the same location:

- The `/etc/cloudera-scm-server/db.properties` file
- The contents of the `/var/lib/cloudera-scm-eventserver` directory

9. On each host in the cluster, restore the `/etc/cloudera-scm-agent/config.ini` file from your backup.

10 Start Cloudera Manager by running the following command on the Cloudera Manager Server host:

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

11 Start the Cloudera Manager Agents by running the following command on all hosts in your clusters:

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

12 Log in to the Cloudera Manager Admin console and start the Cloudera Management Service:

- Select **Clusters > Cloudera Management Service**.

b. Select **Actions** > **Start**.

13 Start your cluster:

a. Select **Clusters** > *Cluster Name*.

b. Select **Actions** > **Start**.

Restoring Databases

Several steps in the rollback procedures require you to restore previously backed-up databases. The steps for backing up and restoring databases differ depending on the database vendor and version you select for your cluster and are beyond the scope of this document.



Important: Restore the databases to their exact state as of when you took the backup. Do not merge in any changes that may have occurred during the subsequent upgrade.

See the following vendor resources for more information:

- **MariaDB 5.5:** <http://mariadb.com/kb/en/mariadb/backup-and-restore-overview/>
- **MySQL 5.5:** <http://dev.mysql.com/doc/refman/5.5/en/backup-and-recovery.html>
- **MySQL 5.6:** <http://dev.mysql.com/doc/refman/5.6/en/backup-and-recovery.html>
- **PostgreSQL 8.4:** <https://www.postgresql.org/docs/8.4/static/backup.html>
- **PostgreSQL 9.2:** <https://www.postgresql.org/docs/9.2/static/backup.html>
- **PostgreSQL 9.3:** <https://www.postgresql.org/docs/9.3/static/backup.html>
- **Oracle 11gR2:** http://docs.oracle.com/cd/E11882_01/backup.112/e10642/toc.htm

Upgrading to Oracle JDK 1.7

Oracle JDK 1.7 is [required](#) for CDH 5. Follow the instructions on this page to upgrade to Java 1.7 *before* you upgrade to CDH 5.

The process for upgrading to Oracle JDK 1.7 varies depending on whether you have a [Cloudera Manager Deployment](#) or an [Unmanaged Deployment](#).



Important: If you are upgrading from JDK 1.6 to JDK 1.7 and you are using AES-256 bit encryption, you must install new encryption policy files. (In a Cloudera Manager deployment, you can automatically install the policy files; for unmanaged deployments, install them manually.) See [Using AES-256 Encryption](#) on page 133.

For both managed and unmanaged deployments, you must also ensure that the Java Truststores are retained during the upgrade. (See [Creating and Using the Truststore](#).)



Note: Oracle JDK 1.7 is deprecated as of Cloudera Manager 5.x and CDH 5.x and support will be removed in version 6.0.

Upgrading to Oracle JDK 1.7 in a Cloudera Manager Deployment

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

1. Upgrade the JDK using one of the following processes:

- Cloudera Manager installation:

1. When upgrading Cloudera Manager, allow Cloudera Manager to upgrade the JDK when you [upgrade Cloudera Manager Agent packages](#).
2. If you have configured TLS for Cloudera Manager, as described in [Level 0: Basic TLS/SSL Configuration](#), copy the `jssecacerts` file from the previous JDK installation to the new JDK installation. For example:

```
cp previous_java_home/jre/lib/security/jssecacerts new_java_home/jre/lib/security
```

(Substitute `previous_java_home` and `new_java_home` with the paths to the JDK installations.)

- Manual installation



1. On each cluster host:

- a. Install the same [supported version](#) of JDK 1.7. See [Java Development Kit Installation](#) for instructions.
- b. Verify that you have set `JAVA_HOME` on each host to the directory where you installed JDK 1.7.
- c. If you have configured TLS for Cloudera Manager, as described in [Level 0: Basic TLS/SSL Configuration](#), copy the `jssecacerts` file from the previous JDK installation to the new JDK installation. For example:

```
cp previous_java_home/jre/lib/security/jssecacerts new_java_home/jre/lib/security
```

(Substitute `previous_java_home` and `new_java_home` with the paths to the JDK installations.)

2. In the Cloudera Manager Admin Console, restart all services:

- a. On the **Home > Status** tab, click  next to the cluster name and select **Restart**.
 - b. In the confirmation dialog box, click **Restart**.
3. If the cluster host is also running Cloudera Management Service roles, restart the Cloudera Management Service.
 - a. Do one of the following:
 - 1. Select **Clusters > Cloudera Management Service**.
 - 2. Select **Actions > Restart**.
 - On the **Home > Status** tab, click  to the right of **Cloudera Management Service** and select **Restart**.
 - b. Click **Restart** to confirm. The **Command Details** window shows the progress of stopping and then starting the roles.
 - c. When **Command completed with n/n successful subcommands** appears, the task is complete. Click **Close**.
 4. Delete your previous Java version files.

Upgrading to Oracle JDK 1.7 in an Unmanaged Deployment



Important:

- Follow these command-line instructions on systems that do not use Cloudera Manager.
- This information applies specifically to CDH 5.11.x. See [Cloudera Documentation](#) for information specific to other releases.

1. Shut down the cluster, following directions in the documentation for the CDH 4 release you are currently running.
2. On each cluster host:
 - a. Install the same [supported version](#) of JDK 1.7. See [Java Development Kit Installation](#) for instructions.
 - b. Verify that you have set `JAVA_HOME` on each host to the directory where you installed JDK 1.7.
 - c. If you have configured TLS for Cloudera Manager, as described in [Level 0: Basic TLS/SSL Configuration](#), copy the `jssecacerts` file from the previous JDK installation to the new JDK installation. For example:

```
cp previous_java_home/jre/lib/security/jssecacerts new_java_home/jre/lib/security
```

(Substitute *previous_java_home* and *new_java_home* with the paths to the JDK installations.)

3. Start the [CDH upgrade](#).
4. Delete your previous Java version files.

Upgrading to Oracle JDK 1.8

Cloudera Manager 5.3 and higher and CDH 5.3 and higher support Oracle JDK 1.8. For other supported versions, see [CDH and Cloudera Manager Supported JDK Versions](#). In some cases, using JDK 1.8 can cause minor performance degradation compared to JDK 1.7.



Warning:

- Cloudera does not support upgrading to JDK 1.8 while upgrading to Cloudera Manager 5.3 or higher. The Cloudera Manager Server must be upgraded to 5.3 or higher before you start.
- Cloudera does not support upgrading to JDK 1.8 while upgrading a cluster to CDH 5.3 or higher. The cluster must be running CDH 5.3 or higher before you start.
- Cloudera does not support a rolling upgrade to JDK 1.8. You must shut down the entire cluster.
- If you are upgrading from a lower major version of the JDK to JDK 1.8 or from JDK 1.6 to JDK 1.7, and you are using AES-256 bit encryption, you must install new encryption policy files. (In a Cloudera Manager deployment, you automatically install the policy files; for unmanaged deployments, install them manually.) See [Using AES-256 Encryption](#) on page 133.

For both managed and unmanaged deployments, you must also ensure that the Java Truststores are retained during the upgrade. (See [Creating and Using the Truststore](#).)

The process for upgrading to Oracle JDK 1.8 varies depending on whether you have a [Cloudera Manager Deployment](#) or an [Unmanaged Deployment](#).

Upgrading to Oracle JDK 1.8 in a Cloudera Manager Deployment

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

1. [Upgrade to Cloudera Manager 5.3 or higher](#) if you have not done so.
2. [Upgrade to CDH 5.3 or higher](#) if you have not done so.
3. Stop the Cloudera Management Service.
4. Stop all clusters.
5. Stop all Cloudera Manager Agents.
6. Stop the Cloudera Manager Server.
7. On the Cloudera Manager Server host and each cluster host:
 - a. Install the same [supported version](#) of JDK 1.8. See [Java Development Kit Installation](#) for instructions.
8. On the Cloudera Manager Server host, configure the location of the JDK in `/etc/default/cloudera-scm-server`.
9. Start the Cloudera Manager Server.
- 10 Start all Cloudera Manager Agents.
- 11 Configure the location of the JDK on cluster hosts as described in [Configuring a Custom Java Home Location](#).
- 12 If you have configured TLS for Cloudera Manager, as described in [Level 0: Basic TLS/SSL Configuration](#), copy the `jssecacerts` file from the previous JDK installation to the new JDK installation. For example:

```
cp previous_java_home/jre/lib/security/jssecacerts new_java_home/jre/lib/security
```

(Substitute *previous_java_home* and *new_java_home* with the paths to the JDK installations.)

- 13 Start all clusters.
- 14 Start the Cloudera Management Service.
- 15 Delete your previous Java version files.

Upgrading to Oracle JDK 1.8 in an Unmanaged Deployment



Important:

- Follow these command-line instructions on systems that do not use Cloudera Manager.
- This information applies specifically to CDH 5.11.x. See [Cloudera Documentation](#) for information specific to other releases.

1. [Upgrade to CDH 5.3 or higher](#) if you have not already done so.
2. [Shut down the cluster.](#)
3. On each cluster host:
 - a. Install the same [supported version](#) of JDK 1.8. See [Java Development Kit Installation](#) for instructions.
 - b. Verify that you have set `JAVA_HOME` on each host to the directory where you installed JDK 1.8 and created a symbolic link to it.
 - c. If you have configured TLS for Cloudera Manager, as described in [Level 0: Basic TLS/SSL Configuration](#), copy the `jssecacerts` file from the previous JDK installation to the new JDK installation. For example:

```
cp previous_java_home/jre/lib/security/jssecacerts new_java_home/jre/lib/security
```

(Substitute *previous_java_home* and *new_java_home* with the paths to the JDK installations.)

4. [Start the cluster.](#)
5. Delete your previous Java version files.

Using AES-256 Encryption

If you are using CentOS/Red Hat Enterprise Linux 5.6 or higher, or Ubuntu, which use AES-256 encryption by default for tickets, you must install the [Java Cryptography Extension \(JCE\) Unlimited Strength Jurisdiction Policy File](#) on all cluster and Hadoop user machines. For JCE Policy File installation instructions, see the `README.txt` file included in the `jce_policy-x.zip` file.

Alternatively, you can configure Kerberos to not use AES-256 by removing `aes256-cts:normal` from the `supported_enctypes` field of the `kdc.conf` or `krb5.conf` file. After changing the `kdc.conf` file, you must restart both the KDC and the `kadmin` server for those changes to take affect. You may also need to re-create or change the password of the relevant principals, including potentially the Ticket Granting Ticket principal (`krbtgt/REALM@REALM`). If AES-256 is still used after completing steps, the `aes256-cts:normal` setting existed when the Kerberos database was created. To fix this, create a new Kerberos database and then restart both the KDC and the `kadmin` server.

To verify the type of encryption used in your cluster:

1. On the local KDC host, type this command to create a test principal:

```
$ kadmin -q "addprinc test"
```

2. On a cluster host, type this command to start a Kerberos session as test:

```
$ kinit test
```

3. On a cluster host, type this command to view the encryption type in use:

```
$ klist -e
```

If AES is being used, output like the following is displayed after you type the `klist` command; note that AES-256 is included in the output:

```
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: test@SCM
Valid starting      Expires              Service principal
05/19/11 13:25:04  05/20/11 13:25:04  krbtgt/SCM@SCM
    Etype (skey, tkt): AES-256 CTS mode with 96-bit SHA-1 HMAC, AES-256 CTS mode with
    96-bit SHA-1 HMAC
```

Upgrading Cloudera Navigator Components

Cloudera Navigator is always upgraded when you upgrade Cloudera Manager.

You can upgrade other Cloudera Navigator components at any time. You do not have to perform these upgrades when upgrading Cloudera Manager or CDH.

See the following topics for detailed upgrade procedures:

Upgrading the Cloudera Navigator Data Management Component

Minimum Required Role: [Navigator Administrator](#) (also provided by **Full Administrator**)

The Cloudera Navigator Data Management component is always upgraded automatically when you upgrade Cloudera Manager.

Before upgrading the Cloudera Navigator management component, review the following version-specific notes that apply to your upgrade:

- 2.9 - Solr indexing has been optimized to improve search speed. No action is required; when Cloudera Navigator is started after an upgrade, the Solr indexing upgrade begins. You will see a message indicating the upgrade progress, and no other action is possible until the upgrade is complete. This could take a number of hours, depending on the amount of data.
- 2.6 and lower - The Cloudera Navigator Metadata Server requires an upgrade of data in the storage directory. Follow the procedure in [Upgrading the Component](#) on page 135.
- 2.4–2.6 - If you are using JDK 1.6, you *must* upgrade to JDK 1.7 or 1.8. See [Java Development Kit Installation](#).
- 2.2 - Policies created with Cloudera Navigator 2.1 (containing the Beta version policy engine) are not retained when upgrading to Cloudera Navigator 2.2.
- 2.1 - When you upgrade to Cloudera Navigator 2.1 or higher from Navigator 2.0 or lower, the upgrade wizard adds a database for the Navigator Metadata Server. This is a different database than the currently existing Navigator Audit Server database.
- 2.0 - Cloudera does not provide an upgrade path from the Navigator Metadata Server that was a beta release in Cloudera Navigator 1.2 to the Cloudera Navigator 2 release. If you are upgrading from Cloudera Navigator 1.2 (included with Cloudera Manager 5.0), you must perform a clean install of Cloudera Navigator 2. If you have Cloudera Navigator roles from a 1.2 release:
 1. Delete the Navigator Metadata Server role.
 2. Remove the contents of the Navigator Metadata Server storage directory.
 3. Add the Navigator Metadata Server role according to the process described in [Adding the Navigator Metadata Server](#).
 4. Clear the cache of any browser that used the 1.2 release of the Navigator Metadata component. Otherwise, you may see errors in the Navigator Metadata UI.

Upgrading the Component



Warning: During Navigator upgrade, a very large Navigator Metadata Server storage directory size can cause out-of-memory (OOM) errors, or can cause upgrade to take a very long time to complete (more than 10 hours). To help prevent OOM errors and decrease upgrade time, run the [purge](#) command before upgrading. If you do not purge, then temporarily increase available JVM memory to 31 GB. Setting JVM memory higher than 31 GB can have a negative impact. Also keep in mind that increasing JVM memory can negatively impact available OS RAM, which can significantly increase upgrade time.

To upgrade the Cloudera Navigator data management component:

Upgrading Cloudera Navigator Components

1. If upgrading from Cloudera Navigator 2.6 or lower, do the following before upgrading:
 - a. Stop the Navigator Metadata Server role.
 - b. Back up the [Navigator Metadata Server storage directory](#).
 - c. [Make sure that the Navigator Metadata Server has sufficient memory](#) to complete the upgrade.
 - d. If you are using an Oracle database, in SQL*Plus, ensure that the following additional privileges are set:

```
GRANT EXECUTE ON sys.dbms_crypto TO nav;  
GRANT CREATE VIEW TO nav;
```

where *nav* is the user of the Navigator Audit Server database.

2. Upgrade Cloudera Manager, following the instructions in [Cloudera Upgrade Overview](#) on page 8. For information on compatible Cloudera Navigator and Cloudera Manager versions, see the [Product Compatibility Matrix for Cloudera Navigator](#) product compatibility matrix.
3. If upgrading from Navigator 2.6 or lower, [start and log into the Cloudera Navigator data management component UI](#). The Upgrading Navigator page displays. Depending on the amount of data in the Navigator Metadata Server storage directory, the upgrade process can take *three to four hours* or longer.
4. When the upgrade is complete, click **Continue**. The Cloudera Navigator landing page is displayed.

Related Information

- [Cloudera Navigator 2 Overview](#)
- [Installing the Cloudera Navigator Data Management Component](#)
- [Upgrading Cloudera Navigator Key Trustee Server](#) on page 136
- [Upgrading Cloudera Navigator Key HSM](#) on page 151
- [Upgrading Cloudera Navigator Encrypt](#) on page 153
- [Cloudera Navigator Data Management Component Administration](#)
- [Cloudera Data Management](#)
- [Configuring Encryption](#)
- [Configuring Authentication in the Cloudera Navigator Data Management Component](#)
- [Configuring TLS/SSL for Cloudera Navigator Metadata Server](#)
- [Cloudera Navigator Data Management Component User Roles](#)

Upgrading Cloudera Navigator Key Trustee Server

Navigator Key Trustee Server 5.4.x is the first release that supports installation using Cloudera Manager. If you are using Cloudera Manager, you must upgrade Key Trustee Server to 5.4 or higher using the command line or the [ktupgrade script](#) before you can migrate Key Trustee Server to Cloudera Manager control.

To upgrade Key Trustee Server from 3.8 to 5.5 or higher, use the [ktupgrade script](#) to simplify the upgrade process.

Upgrading Cloudera Navigator Key Trustee Server 3.x to 5.4.x

Navigator Key Trustee Server 5.4.x is the first release that supports installation using Cloudera Manager. If you are using Cloudera Manager, you must upgrade Key Trustee Server using the command line before you can migrate Key Trustee Server to Cloudera Manager control.

To upgrade Key Trustee Server to 5.5 or higher, see [Upgrading Cloudera Navigator Key Trustee Server 3.8 to 5.5 Using the ktupgrade Script](#) on page 141.



Note: Before upgrading Key Trustee Server, back up the Key Trustee Server database and configuration directory. See [Backing Up Key Trustee Server Manually](#) for instructions.

Upgrading Key Trustee Server 3.x to 5.4.x Using the Command Line

The following instructions apply to both standalone and high availability Key Trustee Servers. For standalone Key Trustee Server, follow the instructions that refer to the *active* Key Trustee Server. For high availability Key Trustee Servers, follow the instructions on all Key Trustee Servers, unless otherwise indicated.

Upgrade Key Trustee Server

1. Stop the httpd service:

```
$ sudo service httpd stop
```

2. Install the EPEL Repository

Dependent packages are available through the Extra Packages for Enterprise Linux (EPEL) repository. To install the EPEL repository, install the `epel-release` package:

1. Copy the URL for the `epel-release-<version>.noarch` file for RHEL 6 or RHEL 7 located in the [How can I use these extra packages?](#) section of the EPEL wiki page.
2. Run the following commands to install the EPEL repository:

```
$ sudo wget <epel_rpm_url>
$ sudo yum install epel-release-<version>.noarch.rpm
```

Replace `<version>` with the version number of the downloaded RPM (for example, 6-8).

If the `epel-release` package is already installed, you see a message similar to the following:

```
Examining /var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: epel-release-6-8.noarch
/var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: does not update installed package.
Error: Nothing to do
```

Confirm that the EPEL repository is installed:

```
$ sudo yum repolist | grep -i epel
```

3. Install the Cloudera Repository

Create or edit the `/etc/yum.repos.d/gazzang.repo` file (for example, `sudo vi /etc/yum.repos.d/gazzang.repo`) and add the following text. Replace `USER` and `PASSWD` with the username and password provided by Cloudera. If you do not know your username or password, contact your Cloudera account team.

```
[gazzang_stable]
name=RHEL $releasever - gazzang.com - base
baseurl=https://USER:PASSWD@archive.gazzang.com/redhat/stable/$releasever
enabled=1
gpgcheck=1
gpgkey=http://archive.gazzang.com/gpg_gazzang.asc
```



Important: If you are using CentOS, add the following line to the CentOS base repository:

```
exclude=python-psycopg2*
```

By default, the base repository is located at `/etc/yum.repos.d/CentOS-Base.repo`. If you have an internal mirror of the base repository, update the correct file for your environment.

Import the GPG key by running the following command:

```
$ sudo rpm --import http://archive.gazzang.com/gpg_gazzang.asc
```

4. Upgrade Key Trustee Server using `yum`:

```
$ sudo yum update keytrustee-server python-keytrustee
```

5. Start the `httpd` service:

```
$ sudo service httpd start
```

Migrate Apache Web Server to CherryPy



Note: Confirm that all ports listed in [Network Requirements](#) are open before proceeding.

For versions 5.4.0 and higher, Key Trustee Server uses CherryPy for the front end web interface; lower versions use the Apache web server. The CherryPy service is managed using the `keytrustee` service. The Apache web server is managed using the `httpd` service. Run the following commands to migrate the web server from Apache to CherryPy.

1. On the active Key Trustee Server, run the `ktadmin db --configure` command as follows:

```
$ sudo -u keytrustee ktadmin db --configure --port 11381 --pg-rootdir  
/var/lib/keytrustee/db --slave keytrustee02.example.com
```

Replace `keytrustee02.example.com` with the hostname of the passive Key Trustee Server. For standalone Key Trustee Server, omit the `--slave keytrustee02.example.com` portion of the command.

2. Export the active Key Trustee Server database. Run the following commands on the active Key Trustee Server:

```
$ sudo -u postgres pg_dump keytrustee > /var/lib/keytrustee/ktdbexport.pgsql  
$ chown keytrustee:keytrustee /var/lib/keytrustee/ktdbexport.pgsql
```

3. Start the Key Trustee Server database and import `ktdbexport.pgsql`:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start --log  
/var/lib/keytrustee/db/pg_ctl.log  
$ sudo -u keytrustee /usr/pgsql-9.3/bin/createdb --host /tmp --port 11381 -O keytrustee  
keytrustee  
$ sudo -u keytrustee psql -d keytrustee -h /tmp -p 11381 <  
/var/lib/keytrustee/ktdbexport.pgsql
```



Note: The `/etc/init.d/postgresql` script does not work when the PostgreSQL database is started by Key Trustee Server, and cannot be used to monitor the status of the database. Use `/etc/init.d/keytrustee-db` instead.

4. (High Availability Key Trustee Servers Only) Start the passive Key Trustee Server. Run the following commands on the passive Key Trustee Server:

```
$ sudo -u keytrustee ktadmin --confdir /var/lib/keytrustee/.keytrustee init-slave --master  
keytrustee01.example.com --pg-rootdir /var/lib/keytrustee/db --no-import-key  
--master-host-port 11381 --logdir /var/lib/keytrustee/.keytrustee/logs  
--postgres-config=local --no-start  
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start
```

5. Edit `/var/lib/keytrustee/.keytrustee/keytrustee.conf` on the active and passive Key Trustee Servers to reference the new database and port. Set the `DB_CONNECT` parameter as follows:

```
"DB_CONNECT": "postgresql://localhost:11381/keytrustee?host=/tmp",
```

6. Restart the Apache web server. Run this command on all Key Trustee Servers:

```
$ sudo service httpd restart
```

7. Start the Key Trustee daemon (which starts the CherryPy web server). Run this command on all Key Trustee Servers:

```
$ sudo /etc/init.d/keytrusteed start
```

8. After verifying that the Key Trustee daemon and CherryPy web server are running, stop the Apache web server and original database and prevent them from starting after reboots. Run these commands on all Key Trustee Servers:

```
$ sudo service httpd stop
$ sudo -u postgres /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/pgsql/9.3/keytrustee stop
$ sudo chkconfig httpd off
$ sudo chkconfig postgresql-9.3 off
```

(High Availability Key Trustee Servers Only) Enable Synchronous Replication

Run the following command on the active Key Trustee Server to enable synchronous replication after upgrading:

```
$ sudo -u keytrustee ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Validating Key Operations

Verify that the upgrade was successful by running the following command on all Key Trustee Servers. The output should be similar to the following. If high availability is enabled, the output should be identical on all Key Trustee Servers:

```
$ curl -k https://keytrustee.example.com:11371/?a=fingerprint
4096R/4EDC46882386C827E20DEEA2D850ACA33BEDB0D1
```

Replace `keytrustee.example.com` with the fully qualified domain name (FQDN) of each Key Trustee Server you are validating.

If you are using Key Trustee Server as the backing key store for [HDFS Transparent Encryption](#), run the following commands to verify that Hadoop key operations are successful:

```
$ hadoop key create hadoop_test_key
$ hadoop key list
$ hadoop key delete hadoop_test_key
```

Migrating Unmanaged Key Trustee Server to Cloudera Manager



Important: If you are upgrading to Key Trustee Server 5.5 or higher without the [ktupgrade script](#), skip this step and continue to [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher](#) on page 146.

For simplified and centralized administration, perform the following steps to move Key Trustee Server under Cloudera Manager control after upgrading Key Trustee Server:

1. Download the Key Trustee Server CSD from the following location:

```
https://archive.gazzang.com/parcels/cloudera/keytrustee-server/5.4.9/stable/latest/csd/
```

When prompted, enter your credentials. If you do not know your credentials, contact your Cloudera account team.

2. Install the CSD into Cloudera Manager as instructed in [Custom Service Descriptor Files](#). The CSD can only be installed on parcel-deployed clusters.

3. Add the following parcel repository to Cloudera Manager following the instructions in [Configuring Cloudera Manager Server Parcel Settings](#):

```
https://<username>:<password>@archive.gazzang.com/parcels/cloudera/keytrustee-server/5.4.9/stable/latest
```

Replace <username> and <password> with your credentials. If you do not know your credentials, contact your Cloudera account team.

4. **(Recommended)** Create a new cluster in Cloudera Manager containing only the hosts the Key Trustee Server will be installed on. Cloudera strongly recommends installing Key Trustee Server in a dedicated cluster to enable multiple clusters to share the same Key Trustee Server and to avoid restarting the Key Trustee Server when restarting a cluster. See [Adding and Deleting Clusters](#) for instructions on how to create a new cluster in Cloudera Manager.
5. Download, distribute, and activate the Key Trustee Server parcel, following the instructions in [Managing Parcels](#). After you activate the Key Trustee Server parcel, Cloudera Manager prompts you to restart the cluster. Click the **Close** button to ignore this prompt. You *do not* need to restart the cluster after installing Key Trustee Server.
6. Stop the active and passive Key Trustee Server web servers using the command that corresponds to your backing web server. See [Migrate Apache Web Server to CherryPy](#) on page 138 for more information.

For Apache web servers:

```
$ sudo service httpd stop
```

For CherryPy web servers:

```
$ sudo service keytrusteed stop
```

7. Stop the active Key Trustee Server database. Run the following command on the active Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db stop
```



Warning: Do not stop the passive Key Trustee Server database. If it is stopped, start it before proceeding by running the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start
```

8. Add the Key Trustee Server service to your cluster, following the instructions in [Adding a Service](#). When customizing role assignments, assign the Active Key Trustee Server and Active Database roles to the active Key Trustee Server host, and the Passive Key Trustee Server and Passive Database roles to the passive Key Trustee Server host.
9. Stop the passive Key Trustee Server database. Run the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db stop
```

- 10 Restart the Key Trustee Server service (**Key Trustee Server service > Actions > Restart**).



Important: Starting or restarting the Key Trustee Server service attempts to start the Active Database and Passive Database roles. If the Active Database is not running when the Passive Database attempts to start, the Passive Database fails to start. If this occurs, manually restart the Passive Database role after confirming that the Active Database role is running.

- 11. (High Availability Key Trustee Servers Only)** Enable synchronous replication. Run the following command on the active Key Trustee Server:

```
$ sudo -u keytrustee ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Updating Key Trustee Server Clients

After upgrading Key Trustee Server to 5.4 or higher, you must configure Key Trustee Server clients (namely Key Trustee KMS and Cloudera Navigator Encrypt) to communicate with Key Trustee Server over the new ports:

- **Key Trustee KMS**

Add the following entries to the Key Trustee KMS advanced configuration snippet (**Key Trustee KMS service > Configuration > Advanced > Key Management Server Advanced Configuration Snippet (Safety Valve) for kms-site.xml**):

```
<property>
  <name>cloudera.trustee.keyprovider.hkpport</name>
  <value>hkp_port_number</value>
  <description>
    Indicates the HTTP port on which Key Trustee Server clients should request public
    keys.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually port
    80 (unencrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    port 11371 (SSL-encrypted).
  </description>
</property>
<property>
  <name>cloudera.trustee.keyprovider.ktsport</name>
  <value>kts_port_number</value>
  <description>
    Indicates the HTTPS port on which the client sends and receives Key Trustee Server
    protocol messages.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually port
    443 (SSL-encrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    port 11371 (SSL-encrypted).
  </description>
</property>
<property>
  <name>cloudera.trustee.keyprovider.hkpssl</name>
  <value>boolean</value>
  <description>
    Indicates whether the client should communicate with the HKP server over an
    SSL-encrypted (true) or unencrypted (false) channel.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually false
    (unencrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    true (SSL-encrypted).
  </description>
</property>
```

- **Cloudera Navigator Encrypt**

See [Updating Key Trustee Server Ports](#) for instructions on updating Cloudera Navigator Encrypt to use the new ports.

Upgrading Cloudera Navigator Key Trustee Server 3.8 to 5.5 Using the ktupgrade Script

Cloudera provides a Python script (`ktupgrade`) to simplify upgrading Key Trustee Server 3.8 to 5.5. The script upgrades package-based Key Trustee Server 3.8 to package-based Key Trustee Server 5.5 and switches the web server from Apache to CherryPy. After the upgrade completes, you must manually migrate Key Trustee Server to use parcels and be managed by Cloudera Manager.

To upgrade from 3.x to 5.5 manually, you must first upgrade to 5.4, and then upgrade to 5.5:

- [Upgrading Cloudera Navigator Key Trustee Server 3.x to 5.4.x](#) on page 136
- [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher](#) on page 146

Prerequisites



Important: The `ktupgrade` script supports upgrading from version 3.8.0 or 3.8.1 to version 5.5.0 or 5.5.2 only. To upgrade to a version higher than 5.5.2, use the `ktupgrade` script to upgrade to 5.5.2, and then follow the instructions in [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher](#) on page 146 to upgrade to the version you want.

- Before upgrading Key Trustee Server, upgrade Cloudera Manager and CDH. See [Cloudera Upgrade Overview](#) on page 8 and [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54. If you are upgrading Key Trustee Server to a version higher than 5.5.2, you can upgrade Cloudera Manager and CDH directly to the version you want before continuing; you do not need to upgrade Cloudera Manager and CDH to 5.5 and complete the Key Trustee Server upgrade before upgrading Cloudera Manager and CDH to a higher version. The Cloudera Manager version must be equal to or higher than the Key Trustee Server version. See [Product Compatibility Matrix for Cloudera Navigator Encryption](#) for more information.
- If you are using [HDFS Transparent Encryption](#) with Key Trustee Server, upgrade Key Trustee KMS. See [Upgrading Key Trustee KMS](#) on page 152 for instructions.
- You must run the `ktupgrade` script as `root`.
- The `ktupgrade` script uses `yum` to upgrade Key Trustee Server. If the Key Trustee Server host does not have Internet access, you must download the Key Trustee Server dependencies from a host with Internet access and copy them to the Key Trustee Server host:

1. Create a temporary directory to store the packages:

```
$ mkdir tmp-keytrustee
```

2. Download the `bigtop-utils` package from the CDH repository:

```
$ sudo wget -P tmp-keytrustee <url>
```

Replace `<url>` with the URL corresponding to the Key Trustee Server version to which you are upgrading:

Table 4: URL for `bigtop-utils` Package

Key Trustee Server Version	URL
5.5.2	http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5.5.2/RPMS/noarch/bigtop-utils-0.7.0+cdh5.5.2+0-1.cdh5.5.2.p0.10.el6.noarch.rpm
5.5.0	http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/5.5.0/RPMS/noarch/bigtop-utils-0.7.0+cdh5.5.0+0-1.cdh5.5.0.p0.15.el6.noarch.rpm

3. Download the `python-paste` and `python-cherrypy` packages:

```
$ sudo yum install yum-downloadonly
$ sudo yum install --downloadonly --downloadaddir=tmp-keytrustee/ python-paste
python-cherrypy
```

4. Copy the packages to the Key Trustee Server host:

```
$ sudo scp tmp-keytrustee/*.rpm <username>@kts01.example.com:/path/to/tmp-keytrustee
```

Replace `kts01.example.com` with the hostname of the active Key Trustee Server, and `/path/to/tmp-keytrustee` with the path to a directory to which you have access.

Download the ktupgrade Script and Repository Tarball

1. Download the ktupgrade script on the active Key Trustee Server host:

```
$ sudo wget http://archive.gazzang.com/keytrustee/ktupgrade
```

If the Key Trustee Server host does not have Internet access, run the command on an Internet-connected host, and then copy the file to the active Key Trustee Server host.

2. Download the repository tarball for Key Trustee Server [5.5.0](#) or [5.5.2](#):
 - a. Select **Packages** from the **SELECT DOWNLOAD TYPE** drop-down menu.
 - b. Select your operating system from the **SELECT AN OS** drop-down menu.
 - c. Click **DOWNLOAD NOW**.
 - d. Copy the downloaded file to the active Key Trustee Server host. Make sure you put the repository tarball and ktupgrade script in the same directory.

Run the ktupgrade Script



Important: You must run the ktupgrade script as the `root` user. By default, the script upgrades the active Key Trustee Server, and then connects to the passive Key Trustee Server host as `root` over SSH (if you are using Key Trustee Server high availability) to upgrade it. You are prompted twice for the `root` password (first to copy the files, and then for the SSH connection).

If your environment does not allow `root` to log in over SSH, contact [Cloudera Support](#) for assistance.

Upgrade the Active Key Trustee Server Using the ktupgrade Script

1. On the active Key Trustee Server host, change to the directory that contains the ktupgrade script and the repository tarball:

```
# cd /path/to/tmp-keytrustee
```

If the host does not have Internet access, make sure that the dependency files you downloaded in [Prerequisites](#) on page 142 are in the same directory as the script and tarball.

2. Make sure the script is executable:

```
# chmod a+x ktupgrade
```

3. Run the ktupgrade script as follows:

```
# ./ktupgrade upgrade-active-kts key-trustee-server-5.5.2-el6.tar.gz
```

Replace `key-trustee-server-5.5.2-el6.tar.gz` with the file name of the repository tarball you downloaded in [Download the ktupgrade Script and Repository Tarball](#) on page 143.

Downgrade Key Trustee Server Using the ktupgrade Script

If you experience any problems upgrading Key Trustee Server, you can use the script to downgrade to your previous version. Run the following command on the active Key Trustee Server:

```
# cd /path/to/tmp-keytrustee
# ./ktupgrade downgrade-active-kts
```

Migrate Key Trustee Server to Cloudera Manager

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

Upgrading Cloudera Navigator Components

Before continuing, you must create an internal repository for the Key Trustee Server parcel. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see [Creating and Using a Parcel Repository for Cloudera Manager](#).

After creating the internal Key Trustee Server parcel repository, do the following:

1. Create a new cluster in Cloudera Manager containing only the Key Trustee Server hosts. This enables multiple clusters to share the same Key Trustee Server and avoids restarting Key Trustee Server when restarting a cluster. See [Adding and Deleting Clusters](#) for instructions on how to create a new cluster in Cloudera Manager.
2. Download, distribute, and activate the Key Trustee Server parcel, following the instructions in [Managing Parcels](#). After you activate the Key Trustee Server parcel, Cloudera Manager prompts you to restart the cluster. Click the **Close** button to ignore this prompt. You *do not* need to restart the cluster after installing Key Trustee Server.
3. Stop the active and passive Key Trustee Server web servers by running the following command on all Key Trustee Server hosts:

```
$ sudo -u keytrustee service keytrusteed stop
```

4. Stop the active Key Trustee Server database by running the following command on the active Key Trustee Server:

```
$ sudo -u keytrustee service keytrustee-db stop
```



Warning: Do not stop the passive Key Trustee Server database. If it is stopped, start it before proceeding by running the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee service keytrustee-db start
```

5. Add the Key Trustee Server service to your cluster, following the instructions in [Adding a Service](#). When customizing role assignments, assign the Active Key Trustee Server and Active Database roles to the active Key Trustee Server host, and the Passive Key Trustee Server and Passive Database roles to the passive Key Trustee Server host.
6. Stop the passive Key Trustee Server database. Run the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee service keytrustee-db stop
```

7. Restart the Key Trustee Server service (**Key Trustee Server service** > **Actions** > **Restart**).



Important: Starting or restarting the Key Trustee Server service attempts to start the Active Database and Passive Database roles. If the Active Database is not running when the Passive Database attempts to start, the Passive Database fails to start. If this occurs, manually restart the Passive Database role after confirming that the Active Database role is running.

8. **(High Availability Key Trustee Servers Only)** Enable synchronous replication. Run the following command on the active Key Trustee Server:

```
$ sudo -u keytrustee ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Validate Key Operations

Verify that the upgrade was successful by running the following command on all Key Trustee Servers. The output should be similar to the following. If high availability is enabled, the output should be identical on all Key Trustee Servers:

```
$ curl -k https://keytrustee.example.com:11371/?a=fingerprint  
4096R/4EDC46882386C827E20DEEA2D850ACA33BEDB0D1
```

Replace `keytrustee.example.com` with the fully qualified domain name (FQDN) of each Key Trustee Server you are validating.

If you are using Key Trustee Server as the backing key store for [HDFS Transparent Encryption](#), run the following commands to verify that Hadoop key operations are successful:

```
$ hadoop key create hadoop_test_key
$ hadoop key list
$ hadoop key delete hadoop_test_key
```

Updating Key Trustee Server Clients

After upgrading Key Trustee Server to 5.4 or higher, you must configure Key Trustee Server clients (namely Key Trustee KMS and Cloudera Navigator Encrypt) to communicate with Key Trustee Server over the new ports:

- **Key Trustee KMS**

Add the following entries to the Key Trustee KMS advanced configuration snippet (**Key Trustee KMS service > Configuration > Advanced > Key Management Server Advanced Configuration Snippet (Safety Valve) for kms-site.xml**):

```
<property>
  <name>cloudera.trustee.keyprovider.hkpport</name>
  <value>hkp_port_number</value>
  <description>
    Indicates the HTTP port on which Key Trustee Server clients should request public
    keys.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually port
    80 (unencrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    port 11371 (SSL-encrypted).
  </description>
</property>
<property>
  <name>cloudera.trustee.keyprovider.ktsport</name>
  <value>kts_port_number</value>
  <description>
    Indicates the HTTPS port on which the client sends and receives Key Trustee Server
    protocol messages.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually port
    443 (SSL-encrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    port 11371 (SSL-encrypted).
  </description>
</property>
<property>
  <name>cloudera.trustee.keyprovider.hkpssl</name>
  <value>boolean</value>
  <description>
    Indicates whether the client should communicate with the HKP server over an
    SSL-encrypted (true) or unencrypted (false) channel.
    On Key Trustee Server 3.8 (Apache webserver-based) servers, this is usually false
    (unencrypted).
    On Key Trustee Server 5.4 and higher (CherryPy-based) servers, this is usually
    true (SSL-encrypted).
  </description>
</property>
```

- **Cloudera Navigator Encrypt**

See [Updating Key Trustee Server Ports](#) for instructions on updating Cloudera Navigator Encrypt to use the new ports.

(Optional) Upgrade to a Higher Release

If you are upgrading Key Trustee Server to a version higher than 5.5.2, continue to [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher](#) on page 146.

Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher

If you are upgrading Key Trustee Server from 3.8 to 5.5 or higher, see [Upgrading Cloudera Navigator Key Trustee Server 3.8 to 5.5 Using the ktupgrade Script](#) on page 141.



Note: Before upgrading Key Trustee Server, back up the Key Trustee Server. See [Backing Up and Restoring Key Trustee Server and Clients](#) for instructions.

Setting Up an Internal Repository

You must create an internal repository to install or upgrade the Cloudera Navigator data encryption components. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see the following topics:

- [Creating and Using a Parcel Repository for Cloudera Manager](#)
- [Creating and Using a Package Repository for Cloudera Manager](#) on page 157

Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using Cloudera Manager

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



Note: These instructions apply to using Cloudera Manager only. To upgrade Key Trustee Server using the command line, skip to the [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using the Command Line \(CherryPy Web Server\)](#) on page 146 or [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using the Command Line \(Apache Web Server\)](#) on page 148 section.

1. Add your internal parcel repository to Cloudera Manager following the instructions in [Configuring Cloudera Manager Server Parcel Settings](#).
2. Download, distribute, and activate the latest Key Trustee Server parcel on the cluster containing the Key Trustee Server host, following the instructions in [Managing Parcels](#).



Important: The KEYTRUSTEE parcel in Cloudera Manager is *not* the Key Trustee Server parcel; it is the Key Trustee KMS parcel. The parcel name for Key Trustee Server is KEYTRUSTEE_SERVER.

After you activate the Key Trustee Server parcel, Cloudera Manager prompts you to restart the cluster. Click the **Close** button to ignore this prompt. You *do not* need to restart the cluster after installing Key Trustee Server.

3. **(High Availability Key Trustee Servers Only)** Enable synchronous replication. On the active Key Trustee Server, run the following command:

```
$ sudo ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using the Command Line (CherryPy Web Server)



Important: Use these instructions only if you have previously [migrated Key Trustee Server](#) to use the CherryPy web server instead of the Apache web server. Otherwise, skip to [Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using the Command Line \(Apache Web Server\)](#) on page 148.

The following instructions apply to both standalone and high availability Key Trustee Servers. For standalone Key Trustee Server, follow the instructions that refer to the *active* Key Trustee Server. For high availability Key Trustee Servers, follow the instructions on all Key Trustee Servers, unless otherwise indicated.

Upgrade Key Trustee Server

1. Stop the keytrusted service:

```
$ sudo service keytrusted stop
```

2. Install the EPEL Repository

Dependent packages are available through the Extra Packages for Enterprise Linux (EPEL) repository. To install the EPEL repository, install the `epel-release` package:

1. Copy the URL for the `epel-release-<version>.noarch` file for RHEL 6 or RHEL 7 located in the [How can I use these extra packages?](#) section of the EPEL wiki page.
2. Run the following commands to install the EPEL repository:

```
$ sudo wget <epel_rpm_url>
$ sudo yum install epel-release-<version>.noarch.rpm
```

Replace `<version>` with the version number of the downloaded RPM (for example, 6-8).

If the `epel-release` package is already installed, you see a message similar to the following:

```
Examining /var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: epel-release-6-8.noarch
/var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: does not update installed package.
Error: Nothing to do
```

Confirm that the EPEL repository is installed:

```
$ sudo yum repolist | grep -i epel
```

3. Install the Cloudera Repository

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

Import the GPG key by running the following command:

```
$ sudo rpm --import http://repo.example.com/path/to/RPM-GPG-KEY-cloudera
```

4. Install the CDH Repository

Key Trustee Server and Key HSM depend on the `bigtop-utils` package, which is included in the CDH repository. For instructions on adding the CDH repository, see [To add the CDH repository](#). To create a local CDH repository, see [Creating a Local Yum Repository](#) for instructions.

5. Upgrade Key Trustee Server:

```
$ sudo yum update keytrustee-server python-keytrustee
```

6. Start the keytrusted service:

```
$ sudo service keytrusted start
```

(High Availability Key Trustee Servers Only) Enable Synchronous Replication

Run the following command on the active Key Trustee Server to enable synchronous replication after upgrading:

```
$ sudo ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Migrate Key Trustee Server to Cloudera Manager

Skip to [Migrating Unmanaged Key Trustee Server to Cloudera Manager](#) on page 150 for instructions on migrating Key Trustee Server to Cloudera Manager control if you have not already done so during a previous upgrade.

Upgrading Cloudera Navigator Key Trustee Server 5.4.x or Higher Using the Command Line (Apache Web Server)



Important: Use these instructions only if you have *not yet* [migrated Key Trustee Server](#) to use the CherryPy web server instead of the Apache web server. The Apache web server is not supported in versions 5.5 and higher.

The following instructions apply to both standalone and high availability Key Trustee Servers. For standalone Key Trustee Server, follow the instructions that refer to the *active* Key Trustee Server. For high availability Key Trustee Servers, follow the instructions on all Key Trustee Servers, unless otherwise indicated.

Migrate Apache Web Server to CherryPy



Note: Confirm that all ports listed in [Network Requirements](#) are open before proceeding.

For versions 5.4.0 and higher, Key Trustee Server uses CherryPy for the front end web interface; lower versions use the Apache web server. The Apache web server is not supported in versions 5.5 and higher. The CherryPy service is managed using the `keytrusteed` service. The Apache web server is managed using the `httpd` service. Before upgrading, run the following commands to migrate the web server from Apache to CherryPy.

1. On the active Key Trustee Server, run the `ktadmin db --configure` command as follows:

```
$ sudo ktadmin db --configure --port 11381 --pg-rootdir /var/lib/keytrustee/db --slave  
keytrustee02.example.com
```

Replace `keytrustee02.example.com` with the hostname of the passive Key Trustee Server. For standalone Key Trustee Server, omit the `--slave keytrustee02.example.com` portion of the command.

If you use a database directory other than `/var/lib/keytrustee/db`, create or edit the `/etc/sysconfig/keytrustee-db` file and add the following:

```
ARGS="--pg-rootdir /path/to/db"
```

2. Export the Key Trustee Server database. Run the following commands on the active Key Trustee Server:

```
$ sudo -u postgres pg_dump keytrustee > /var/lib/keytrustee/ktdbexport.pgsql  
$ chown keytrustee:keytrustee /var/lib/keytrustee/ktdbexport.pgsql
```

3. Start the Key Trustee Server database and import `ktdbexport.pgsql`:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start --log  
/var/lib/keytrustee/db/pg_ctl.log  
$ sudo -u keytrustee /usr/pgsql-9.3/bin/createdb --host /tmp --port 11381 -O keytrustee  
keytrustee  
$ sudo -u keytrustee psql -d keytrustee -h /tmp -p 11381 <  
/var/lib/keytrustee/ktdbexport.pgsql
```



Note: The `/etc/init.d/postgresql` script does not work when the PostgreSQL database is started by Key Trustee Server, and cannot be used to monitor the status of the database. Use `/etc/init.d/keytrustee-db` instead.

- 4. (High Availability Key Trustee Servers Only)** Start the passive Key Trustee Server. Run the following commands on the passive Key Trustee Server:

```
$ sudo ktadmin --confdir /var/lib/keytrustee/.keytrustee init-slave --master
keytrustee01.example.com --pg-rootdir /var/lib/keytrustee/db --no-import-key
--master-host-port 11381 --logdir /var/lib/keytrustee/.keytrustee/logs
--postgres-config=local --no-start
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start
```

If you use a database directory other than `/var/lib/keytrustee/db`, create or edit the `/etc/sysconfig/keytrustee-db` file and add the following:

```
ARGS="--pg-rootdir /path/to/db"
```

- 5.** Edit `/var/lib/keytrustee/.keytrustee/keytrustee.conf` on all Key Trustee Servers to reference the new database and port. Set the `DB_CONNECT` parameter as follows:

```
"DB_CONNECT": "postgresql://localhost:11381/keytrustee?host=/tmp",
```

- 6.** Restart the Apache web server. Run this command on all Key Trustee Servers:

```
$ sudo service httpd restart
```

- 7.** Start the Key Trustee daemon (which starts the CherryPy web server). Run this command on all Key Trustee Servers:

```
$ sudo service keytrusteed start
```

- 8.** After verifying that the Key Trustee daemon and CherryPy web server are running, stop the Apache web server and original database and prevent them from starting after reboots. Run these commands on all Key Trustee Servers:

```
$ sudo service httpd stop
$ sudo -u postgres /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/pgsql/9.3/keytrustee stop
$ sudo chkconfig httpd off
$ sudo chkconfig postgresql-9.3 off
```

Upgrade Key Trustee Server

- 1.** Stop the `httpd` service:

```
$ sudo service httpd stop
```

- 2. Install the EPEL Repository**

Dependent packages are available through the Extra Packages for Enterprise Linux (EPEL) repository. To install the EPEL repository, install the `epel-release` package:

- Copy the URL for the `epel-release-<version>.noarch` file for RHEL 6 or RHEL 7 located in the [How can I use these extra packages?](#) section of the EPEL wiki page.
- Run the following commands to install the EPEL repository:

```
$ sudo wget <epel_rpm_url>
$ sudo yum install epel-release-<version>.noarch.rpm
```

Replace `<version>` with the version number of the downloaded RPM (for example, 6-8).

If the `epel-release` package is already installed, you see a message similar to the following:

```
Examining /var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: epel-release-6-8.noarch
/var/tmp/yum-root-jmZhL0/epel-release-6-8.noarch.rpm: does not update installed package.
Error: Nothing to do
```

Confirm that the EPEL repository is installed:

```
$ sudo yum repolist | grep -i epel
```

3. Install the Cloudera Repository

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

Import the GPG key by running the following command:

```
$ sudo rpm --import http://repo.example.com/path/to/RPM-GPG-KEY-cloudera
```

4. Upgrade Key Trustee Server:

```
$ sudo yum update keytrustee-server python-keytrustee
```

5. Start the httpd service:

```
$ sudo service httpd start
```

(High Availability Key Trustee Servers Only) Enable Synchronous Replication

Run the following command on the active Key Trustee Server to enable synchronous replication after upgrading:

```
$ sudo ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Migrate Key Trustee Server to Cloudera Manager

Continue to [Migrating Unmanaged Key Trustee Server to Cloudera Manager](#) on page 150 for instructions on migrating Key Trustee Server to Cloudera Manager control if you have not already done so during a previous upgrade.

Migrating Unmanaged Key Trustee Server to Cloudera Manager

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

For simplified and centralized administration, perform the following steps to move Key Trustee Server under Cloudera Manager control (if you have not already done so) after upgrading Key Trustee Server:

1. **(Recommended)** Create a new cluster in Cloudera Manager containing only the hosts the Key Trustee Server will be installed on. Cloudera strongly recommends installing Key Trustee Server in a dedicated cluster to enable multiple clusters to share the same Key Trustee Server and to avoid restarting the Key Trustee Server when restarting a cluster. See [Adding and Deleting Clusters](#) for instructions on how to create a new cluster in Cloudera Manager.
2. Download, distribute, and activate the Key Trustee Server parcel, following the instructions in [Managing Parcels](#). After you activate the Key Trustee Server parcel, Cloudera Manager prompts you to restart the cluster. Click the **Close** button to ignore this prompt. You *do not* need to restart the cluster after installing Key Trustee Server.
3. Stop the active and passive Key Trustee Server web servers using the command that corresponds to your backing web server. See [Migrate Apache Web Server to CherryPy](#) on page 148 for more information.

For Apache web servers:

```
$ sudo service httpd stop
```

For CherryPy web servers:

```
$ sudo service keytrusteed stop
```

4. Stop the active Key Trustee Server database. Run the following command on the active Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db stop
```



Warning: Do not stop the passive Key Trustee Server database. If it is stopped, start it before proceeding by running the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db start
```

5. Add the Key Trustee Server service to your cluster, following the instructions in [Adding a Service](#). When customizing role assignments, assign the Active Key Trustee Server and Active Database roles to the active Key Trustee Server host, and the Passive Key Trustee Server and Passive Database roles to the passive Key Trustee Server host.
6. Stop the passive Key Trustee Server database. Run the following command on the passive Key Trustee Server:

```
$ sudo -u keytrustee /usr/pgsql-9.3/bin/pg_ctl -D /var/lib/keytrustee/db stop
```

7. Restart the Key Trustee Server service (**Key Trustee Server service > Actions > Restart**).



Important: Starting or restarting the Key Trustee Server service attempts to start the Active Database and Passive Database roles. If the Active Database is not running when the Passive Database attempts to start, the Passive Database fails to start. If this occurs, manually restart the Passive Database role after confirming that the Active Database role is running.

8. **(High Availability Key Trustee Servers Only)** Enable synchronous replication. Run the following command on the active Key Trustee Server:

```
$ sudo ktadmin enable-synchronous-replication --pg-rootdir /var/lib/keytrustee/db
```

Upgrading Cloudera Navigator Key HSM

Setting Up an Internal Repository

You must create an internal repository to install or upgrade Cloudera Navigator Key HSM. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see [Creating and Using a Package Repository for Cloudera Manager](#) on page 157.

Upgrading Key HSM



Important: If you have implemented Key Trustee Server high availability, upgrade Key HSM on each Key Trustee Server.

1. **Install the Cloudera Repository**

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

Upgrading Cloudera Navigator Components

Import the GPG key by running the following command:

```
$ sudo rpm --import http://repo.example.com/path/to/RPM-GPG-KEY-cloudera
```

2. Install the CDH Repository

Key Trustee Server and Key HSM depend on the `bigtop-utils` package, which is included in the CDH repository. For instructions on adding the CDH repository, see [To add the CDH repository](#). To create a local CDH repository, see [Creating a Local Yum Repository](#) for instructions.

3. Stop the Key HSM Service

Stop the Key HSM service before upgrading:

```
$ sudo service keyhsm shutdown
```

4. Upgrade Navigator Key HSM

Upgrade the Navigator Key HSM package using `yum`:

```
$ sudo yum update keytrustee-keyhsm
```

Cloudera Navigator Key HSM is installed to the `/usr/share/keytrustee-server-keyhsm` directory by default.

5. Start the Key HSM Service

Start the Key HSM service:

```
$ sudo service keyhsm start
```

Upgrading Key Trustee KMS



Important: Following these instructions upgrades the software for the Key Trustee KMS service; this enables you to use Cloudera Navigator Key Trustee Server as the underlying keystore for [HDFS Transparent Encryption](#). This *does not* upgrade Key Trustee Server. See [Upgrading Cloudera Navigator Key Trustee Server](#) on page 136 for instructions on upgrading Key Trustee Server.

Key Trustee KMS is supported only in Cloudera Manager deployments. You can install the software using parcels or packages, but running Key Trustee KMS outside of Cloudera Manager is not supported.

Setting Up an Internal Repository

You must create an internal repository to upgrade Key Trustee KMS. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see [Creating and Using a Parcel Repository for Cloudera Manager](#) if you are using parcels, or [Creating and Using a Package Repository for Cloudera Manager](#) on page 157 if you are using packages.

Upgrading Key Trustee KMS Using Parcels



Important: Back up Key Trustee KMS before upgrading. See [Backing Up and Restoring Key Trustee Server and Clients](#) for instructions.

1. Go to **Hosts > Parcels**.
2. Click **Configuration** and add your internal repository to the **Remote Parcel Repository URLs** section. See [Configuring the Cloudera Manager Server to Use the Parcel URL for Hosted Repositories](#) for more information.
3. Click **Save Changes**.

- Download, distribute, and activate the KEYTRUSTEE parcel for the version to which you are upgrading. See [Parcels](#) for detailed instructions on using parcels to install or upgrade components.
- Restart the Key Trustee KMS service (**Key Trustee KMS service** > **Actions** > **Restart**).

Upgrading Key Trustee KMS Using Packages

- After [Setting Up an Internal Repository](#) on page 152, configure the Key Trustee KMS host to use the repository. See [Modifying Clients to Find the Repository](#) on page 158 for more information.
- Add the CDH repository. See [To add the CDH repository](#) for instructions. If you want to create an internal CDH repository, see [Creating a Local Yum Repository](#).
- Upgrade the `keytrustee-keyprovider` package using the appropriate command for your operating system:

- **RHEL-compatible**

```
$ sudo yum install keytrustee-keyprovider
```

- **SLES**

```
$ sudo zypper install keytrustee-keyprovider
```

- **Ubuntu or Debian**

```
$ sudo apt-get install keytrustee-keyprovider
```

- Restart the Key Trustee KMS service (**Key Trustee KMS service** > **Actions** > **Restart**).

Upgrading Cloudera Navigator Encrypt

Setting Up an Internal Repository

You must create an internal repository to install or upgrade the Cloudera Navigator data encryption components. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see the following topics:

- [Creating and Using a Parcel Repository for Cloudera Manager](#)
- [Creating and Using a Package Repository for Cloudera Manager](#) on page 157

Upgrading Navigator Encrypt (RHEL-Compatible)



Important: Cloudera supports RHEL 7 with the following limitations:

- Only RHEL 7.2 and 7.1 are supported. RHEL 7.0 is not supported.
- RHEL 7.1 is only supported with CDH 5.5 and higher.
- RHEL 7.2 is only supported with CDH 5.7 and higher.
- Only new installations of RHEL 7.2 and 7.1 are supported by Cloudera. For upgrades to RHEL 7.1 or 7.2, contact your OS vendor and see [Does Red Hat support upgrades between major versions of Red Hat Enterprise Linux?](#)

1. Install the Cloudera Repository

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

Import the GPG key by running the following command:

```
$ sudo rpm --import http://repo.example.com/path/to/gpg_gazzang.asc
```

2. Stop Navigator Encrypt

Stop the Navigator Encrypt service:

```
$ sudo service navencrypt-mount stop
```

For RHEL 7, use `systemctl` instead:

```
$ sudo systemctl stop navencrypt-mount
```

3. Upgrade Navigator Encrypt

Upgrade the Navigator Encrypt client using `yum`:

```
$ sudo yum update navencrypt
```

4. Start Navigator Encrypt

Start the Navigator Encrypt service:

```
$ sudo service navencrypt-mount start
```

For RHEL 7, use `systemctl` instead:

```
$ sudo systemctl start navencrypt-mount
```

Upgrading Navigator Encrypt (SLES)

1. Install the Cloudera Repository

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

Import the GPG key by running the following command:

```
$ sudo rpm --import http://repo.example.com/path/to/gpg_gazzang.asc
```

2. Stop Navigator Encrypt

Stop the Navigator Encrypt service:

```
$ sudo service navencrypt-mount stop
```

3. Upgrade the Navigator Encrypt Client

Upgrade Navigator Encrypt:

```
$ sudo zypper update navencrypt
```

4. Enable Unsupported Modules

Edit `/etc/modprobe.d/unsupported-modules` and set `allow_unsupported_modules` to 1. For example:

```
#
# Every kernel module has a flag 'supported'. If this flag is not set loading
# this module will taint your kernel. You will not get much help with a kernel
# problem if your kernel is marked as tainted. In this case you firstly have
# to avoid loading of unsupported modules.
#
# Setting allow_unsupported_modules 1 enables loading of unsupported modules
# by modprobe, setting allow_unsupported_modules 0 disables it. This can
# be overridden using the --allow-unsupported-modules command line switch.
allow_unsupported_modules 1
```

5. Start Navigator Encrypt

Start the Navigator Encrypt service:

```
$ sudo service navencrypt-mount start
```

Upgrading Navigator Encrypt (Debian or Ubuntu)**1. Install the Cloudera Repository**

Add the internal repository you created. See [Modifying Clients to Find the Repository](#) on page 158 for more information.

- **Ubuntu**

```
$ echo "deb http://repo.example.com/path/to/ubuntu/stable $DISTRIB_CODENAME main" | sudo tee -a /etc/apt/sources.list
```

- **Debian**

```
$ echo "deb http://repo.example.com/path/to/debian/stable $DISTRIB_CODENAME main" | sudo tee -a /etc/apt/sources.list
```

Import the GPG key by running the following command:

```
$ wget -O - http://repo.example.com/path/to/gpg_gazzang.asc | apt-key add -
```

Update the repository index with `apt-get update`.

2. Stop Navigator Encrypt

Stop the Navigator Encrypt service:

```
$ sudo service navencrypt-mount stop
```

3. Upgrade the Navigator Encrypt Client

Upgrade Navigator Encrypt:

```
$ sudo apt-get install navencrypt
```

4. Start Navigator Encrypt

Start the Navigator Encrypt service:

```
$ sudo service navencrypt-mount start
```

Best Practices for Upgrading Navigator Encrypt Hosts

The following lists best practices for upgrading operating systems (OS) and kernels on hosts that have Navigator Encrypt installed:

- Make sure that the version you are upgrading to is supported by Navigator Encrypt. See the product compatibility matrix for [Product Compatibility Matrix for Cloudera Navigator Encryption](#) for more information.
- Always test upgrades in a development or testing environment before upgrading production hosts.
- If possible, upgrade the entire operating system instead of only upgrading the kernel.
- If you need to upgrade the kernel only, make sure that your OS version supports the kernel version to which you are upgrading.

Upgrading Cloudera Navigator Components

- Always back up the `/etc/navencrypt` directory before upgrading. If you have problems accessing encrypted data after upgrading the OS or kernel, restore `/etc/navencrypt` from your backup and try again.

Creating and Using a Package Repository for Cloudera Manager

This topic describes how to create a remote package repository and direct hosts in your Cloudera Manager deployment to use that repository. There are two options for publishing the repository:

- [Creating a Permanent Remote Repository](#) on page 157
- [Creating a Temporary Remote Repository](#) on page 158

Once you have created a repository, go to [Modifying Clients to Find the Repository](#) on page 158.

After completing these steps, you have established the environment required to install a previous version of Cloudera Manager or install Cloudera Manager to hosts that are not connected to the Internet. Proceed with the installation process, being sure to target the newly created repository with your package management tool.

Creating a Permanent Remote Repository

Installing a Web Server

The repository is typically hosted using HTTP on a host inside your network. If you already have a web server in your organization, you can move the repository directory, which will include both the RPMs and the `repodata/` subdirectory, to some a location hosted by the web server. An easy web server to install is the Apache HTTPD. If you are able to use an existing web server, then note the URL and skip to [Downloading the Tarball and Publishing Repository Files](#) on page 157.

Installing Apache HTTPD

You may need to respond to some prompts to confirm you want to complete the installation.

OS	Command
RHEL	<code>[root@localhost yum.repos.d]\$ yum install httpd</code>
SLES	<code>[root@localhost zypp]\$ zypper install httpd</code>
Ubuntu or Debian	<code>[root@localhost apt]\$ apt-get install httpd</code>

Starting Apache HTTPD

OS	Command
RHEL	<code>[root@localhost tmp]\$ service httpd start</code>
SLES	<code>[root@localhost tmp]\$ service apache2 start</code>
Ubuntu or Debian	<code>[root@localhost tmp]\$ service apache2 start</code>

Downloading the Tarball and Publishing Repository Files

1. Download the tarball for your OS distribution from the [repo as tarball archive](#).

For Cloudera Navigator data encryption components, go to the download page for each component, select your OS version, and click **Download**:

- [Cloudera Navigator Key Trustee Server](#)
- [Cloudera Navigator Key HSM](#)
- [Cloudera Navigator Key Trustee KMS](#)
- [Cloudera Navigator Encrypt](#)

Creating and Using a Package Repository for Cloudera Manager

2. Unpack the tarball, move the files to the web server directory, and modify file permissions. For example, you might use the following commands:

```
[root@localhost tmp]$ tar xvfz cm5.0.0-centos6.tar.gz
[root@localhost tmp]$ mv cm /var/www/html
[root@localhost tmp]$ chmod -R ugo+rX /var/www/html/cm
```

After moving files and changing permissions, visit `http://hostname:port/cm` to verify that you see an index of files. Apache may have been configured to not show indexes, which is also acceptable.

Creating a Temporary Remote Repository

You can quickly create a temporary remote repository to deploy a package once. It is convenient to perform this on the same host that runs Cloudera Manager, or a gateway role. In this example, [python SimpleHTTPServer](#) is used from a directory of your choosing.

1. Download the tarball for your OS distribution from the [repo as tarball archive](#).

For Cloudera Navigator data encryption components, go to the download page for each component, select your OS version, and click **Download**:

- [Cloudera Navigator Key Trustee Server](#)
- [Cloudera Navigator Key HSM](#)
- [Cloudera Navigator Key Trustee KMS](#)
- [Cloudera Navigator Encrypt](#)

2. Unpack the tarball and modify file permissions. For example, you might use the following commands:

```
[root@localhost tmp]$ tar xvfz cm5.0.0-centos6.tar.gz
[root@localhost tmp]$ chmod -R ugo+rX /tmp/cm
```

3. Determine a port that your system is not listening on (for example, port 8900).
4. Change to the directory containing the files.

```
$ cd /tmp/cm
```

5. Start a python SimpleHTTPServer to host these two files:

```
$ python -m SimpleHTTPServer 8900
Serving HTTP on 0.0.0.0 port 8900 ...
```

6. Confirm you can get to this hosted package directory by going to `http://server:8900/cm` in your browser. You should see links for the hosted files.

Modifying Clients to Find the Repository

Having established the repository, modify the clients so they find the repository.

OS	Command
RHEL	Create files on client systems with the following information and format, where <i>hostname</i> is the name of the web server: [myrepo] name=myrepo baseurl=http://hostname/cm/5 enabled=1 gpgcheck=0

OS	Command
	See <code>man yum.conf</code> for more details. Put that file into <code>/etc/yum.repos.d/myrepo.repo</code> on all of your hosts to enable them to find the packages that you are hosting.
SLES	Use the <code>zypper</code> utility to update client system repo information by issuing the following command: <pre>\$ zypper addrepo http://hostname/cm alias</pre>
Ubuntu or Debian	Add a new <code>.list</code> file to <code>/etc/apt/sources.list.d/</code> on client systems. For example, you might create the file <code>/etc/apt/sources.list.d/my-private-cloudera-repo.list</code> . In that file, create an entry to your newly created repository. For example: <pre>\$ cat /etc/apt/sources.list.d/my-private-cloudera-repo.list deb http://hostname/cm codename components</pre> <p>You can find the <i>codename</i> and <i>component</i> variables in the <code>./conf/distributions</code> file in the repository.</p> <p>After adding your <code>.list</code> file, ensure <code>apt-get</code> uses the latest information by issuing the following command:</p> <pre>\$ sudo apt-get update</pre>

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 the Hive Metastore Server role type. This role manages the metastore process when Hive is configured with a remote metastore.

When upgrading from Cloudera Manager versions lower than 4.5, Cloudera Manager automatically creates new Hive services to capture the previous implicit Hive dependency from Hue and Impala. Your previous services continue to function without impact. If Hue was using a Hive metastore backed by a Derby database, the newly created Hive Metastore Server also uses Derby. Because Derby does not allow concurrent connections, Hue continues to work, but the new Hive Metastore Server does not run. The failure is harmless (because nothing uses this new Hive Metastore Server at this point) and intentional, to preserve cluster functionality as it existed before upgrade. Cloudera recommends switching to a different supported database because of the limitations of a Derby-backed Hive metastore.

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

Backing up Databases

Before beginning the upgrade process, shut down the services that are using databases. This includes the Cloudera Manager Management Service roles, the Hive Metastore Server, and Cloudera Navigator, if it is in use. Cloudera strongly recommends that you then back up all databases, however backing up the Activity Monitor database is optional. For information on backing up databases see [Backing Up Databases](#).

Creating New Databases

If any additional databases will be required as a result of the upgrade, complete any required preparatory work to install and configure those databases. The upgrade instructions assume all required databases have been prepared. For more information on required databases, see [Cloudera Manager and Managed Service Datastores](#).

Modifying Databases to Support UTF-8

Cloudera Manager 4 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 username 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.

Modifying 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 five databases on one host (the databases for Cloudera Manager Server, Activity Monitor, Reports Manager, Cloudera Navigator, and Hive metastore), set the maximum connections to 550.

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 service that requires a database and then add 50 extra connections. For example, for two services, set the maximum connections to 250. If you have five services that require a database on one host (the databases for Cloudera Manager Server, Activity Monitor, Reports Manager, Cloudera Navigator, and Hive metastore), set the maximum connections to 550.

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 a database for two services, anticipate 250 maximum connections. If you anticipate a maximum of 250 connections, plan for 280 sessions.

Database Considerations for Cloudera Manager Upgrades

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 [Upgrading Cloudera Manager 4 to Cloudera Manager 5](#) on page 32 or [Upgrading Cloudera Manager 5 Using Packages](#) on page 12.

Re-Running the Cloudera Manager Upgrade Wizard

Minimum Required Role: [Full Administrator](#)

The first time you log in to the Cloudera Manager server after upgrading your Cloudera Manager software, the upgrade wizard runs. If you did not complete the wizard at that time, or if you had hosts that were unavailable at that time and still need to be upgraded, you can re-run the upgrade wizard:

1. Click the **Hosts** tab.
2. Click **Re-run Upgrade Wizard**. This takes you back through the installation wizard to upgrade Cloudera Manager Agents on your hosts as necessary.
3. Select the release of the Cloudera Manager Agent to install. Normally, this is the **Matched Release for this Cloudera Manager Server**. However, if you used a custom repository (instead of `archive.cloudera.com`) for the Cloudera Manager server, select **Custom Repository** and provide the required information. The custom repository allows you to use an alternative location, but that location must contain the matched Agent version.
4. Specify credentials and initiate Agent installation:
 - Select **root** or enter the username for an account that has password-less sudo permission.
 - Select an authentication method:
 - If you choose password authentication, enter and confirm the password.
 - If you choose public-key authentication, provide a passphrase and path to the required key files.
 - You can 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.

When you click **Continue** the Cloudera Manager Agent is upgraded on all the currently managed hosts. You cannot search for new hosts through this process. To add hosts to your cluster, click the **Add New Hosts to Cluster** button.

Reverting a Failed Cloudera Manager Upgrade

If you have a CDH 3 cluster running under Cloudera Manager 4, you cannot upgrade to Cloudera Manager 5 because it does not support CDH 3. Likewise, an upgrade from Cloudera Manager 3 to Cloudera Manager 5 is not supported. In either case, the Cloudera Manager 5 server will not start, and you must now downgrade your Cloudera Manager server, back to the version you were using prior to attempting the upgrade.



Important: The following instructions assume that a Cloudera Manager upgrade failed, and that the upgraded server never started, so that the remaining steps of the upgrade process were not performed. The steps below are not sufficient to revert from a running Cloudera Manager 5 deployment.

Reinstall the Cloudera Manager Server Packages

In this step, you install the Cloudera Manager Server packages to the version you were running previously. You must reinstall the same version of Cloudera Manager you were using previously, so that the version of your Cloudera Manager Agents match the server.

The steps below assume that the Cloudera Manager Server is already stopped (as it failed to start after the attempted upgrade).

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

- RHEL-compatible 7 and higher:

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

- All other Linux distributions:

```
sudo service cloudera-scm-server-db fast_stop
```

2. Reinstall the same Cloudera Manager Server version that you were previously running. You can reinstall from the Cloudera repository at <https://archive.cloudera.com/cm4/> or <https://archive.cloudera.com/cm5/> or alternately, you can create your own repository, as described in [Understanding Custom Installation Solutions](#).

- a. Find the Cloudera repo file for your distribution by starting at <https://archive.cloudera.com/cm4/> or <https://archive.cloudera.com/cm5/> and navigating to the directory that matches your operating system.

For example, for RHEL or CentOS 6, you would go to https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/. Within that directory, find the repo file that contains information including the repository's base URL and GPG key. On CentOS 6, the contents of the `cloudera-manager.repo` file might appear as follows:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 5, on RHEL or CentOS 6 x86_64
name=Cloudera Manager
baseurl=https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/5/
gpgkey = https://archive.cloudera.com/cm5/redhat/6/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,

<https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm><https://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 7.0 x86_64
deb https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
deb-src https://archive.cloudera.com/cm5/debian/wheezy/amd64/cm wheezy-cm5 contrib
```

You must edit the file if it exist and modify the URL to reflect the exact version of Cloudera Manager you are using (unless you want the downgrade to also upgrade to the latest version of Cloudera Manager 4). The possible versions are shown in the directory on archive. Setting the URL (an example):

OS	Command
RHEL	Replace baseurl=https://archive.cloudera.com/cm5/redhat/5/x86_64/cm/5/ with baseurl=https://archive.cloudera.com/cm5/redhat/5/x86_64/cm/5.0.5/
Ubuntu or Debian	Replacedeb https://archive.cloudera.com/cm5/debian/squeeze/amd64/cm squeeze-cm5 contrib with deb https://archive.cloudera.com/cm5/debian/squeeze/amd64/cm squeeze-cm5.0.5 contrib

- b. Copy the repo file to the configuration location for the package management software for your system:

Operating System	Commands
RHEL	Copy <code>cloudera-manager.repo</code> to <code>/etc/yum.repos.d/</code> .
SLES	Copy <code>cloudera-manager.repo</code> to <code>/etc/zypp/repos.d/</code> .
Ubuntu or Debian	Copy <code>cloudera.list</code> to <code>/etc/apt/sources.list.d/</code> .

- c. Run the following commands:

Operating System	Commands
RHEL	<code>\$ sudo yum downgrade 'cloudera-*'</code>
SLES	<code>\$ sudo zypper clean --all</code> <code>\$ sudo zypper dup -r https://archive.cloudera.com/cm4/sles/11/x86_64/cm/4/</code> To download from your own repository: <code>\$ sudo zypper clean --all</code> <code>\$ sudo zypper dup -r http://myhost.example.com/path_to_cm_repo</code>
Ubuntu or Debian	There's no action that will downgrade to the version currently in the repository. Read DowngradeHowto , download the script described therein, run it, and then run <code>apt-get install</code> for the name=version pairs that it provides for Cloudera Manager.

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

```
$ rpm -qa 'cloudera-manager-*'
cloudera-manager-daemons-5.0.5-1.cm505.p0.163.el6.x86_64
cloudera-manager-server-5.0.5-1.cm505.p0.163.el6.x86_64
cloudera-manager-agent-5.0.5-1.cm505.p0.163.el6.x86_64
```

Reverting a Failed Cloudera Manager Upgrade

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-agent 5.0.5-1.cm505.p0.163~sq The Cloudera Manager Agent  
ii cloudera-manager-daemo 5.0.5-1.cm505.p0.163~sq Provides daemons for monitoring Hadoop  
and related tools.  
ii cloudera-manager-serve 5.0.5-1.cm505.p0.163~sq The Cloudera Manager Server
```

You may also see an entry for the `cloudera-manager-server-db` if you are using the embedded database, and additional packages for plug-ins, 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.

Start the Server

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

1. If you are using the embedded PostgreSQL database for Cloudera Manager, start the database:

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

2. Start the server:

```
$ 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.

Upgrading Unmanaged CDH Using the Command Line

This section provides instructions for upgrading CDH to the latest release, using the command line instead of Cloudera Manager. (A cluster you are not managing with Cloudera Manager is referred to as an **unmanaged** cluster.)



Important:

- Follow these command-line instructions on systems that do not use Cloudera Manager.
- This information applies specifically to CDH 5.11.x. See [Cloudera Documentation](#) for information specific to other releases.

To proceed with the upgrade, choose one of the following sections:

Upgrading from CDH 4 to CDH 5



Note: If you are using Cloudera Manager to manage CDH, *do not* use the instructions in this section.

- If you are running Cloudera Manager 4, you must first upgrade to Cloudera Manager 5 because Cloudera Manager 4 cannot manage CDH 5; see [Cloudera Upgrade Overview](#) on page 8.
- Follow directions in [Upgrading CDH 4 to CDH 5](#) on page 83 to upgrade CDH 4 to CDH 5 in a Cloudera Manager deployment.

Use the following information and instructions to upgrade to the latest CDH 5 release from a CDH 4 release:

- [Before You Begin Upgrading to CDH 5 Using the Command Line](#) on page 167
- [Upgrading to CDH 5](#) on page 169



Important: This upgrade requires uninstalling the CDH 4 packages and installing the CDH 5 packages.



Note:

If you are migrating from MapReduce v1 (MRv1) to MapReduce v2 (MRv2, YARN), see [Migrating from MapReduce \(MRv1\) to MapReduce \(MRv2\)](#) for important information and instructions.

Before You Begin Upgrading to CDH 5 Using the Command Line

Before upgrading, be sure to read about the latest [Incompatible Changes and Limitations](#), [Install and Upgrade Known Issues](#) and [Known Issues in CDH 5](#) in the [CDH 5 Release Notes](#). If you are currently running MRv1, you should read [CDH 5 and MapReduce](#) before proceeding.

Plan Downtime

If you are upgrading a cluster that is part of a production system, reserve a maintenance window with enough extra time allotted in case of complications. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.

Upgrading Unmanaged CDH Using the Command Line

Install Java 1.7

CDH 5 [requires Java 1.7](#) or higher. See [Upgrading to Oracle JDK 1.7](#) on page 130, and make sure you have read [Install and Upgrade Known Issues](#) before you proceed with the upgrade.

Delete Symbolic Links in HDFS

If there are symbolic links in HDFS when you upgrade from CDH 4 to CDH 5, the upgrade will fail and you must downgrade to CDH 4, delete the symbolic links, and start over.

To check for symbolic links in CDH 4 HDFS:

1. Go to the directory on the NameNode that contains the latest `fsimage`. The location of this directory is specified as the value of `dfs.namenode.name.dir` (or `dfs.name.dir`) in `hdfs-site.xml`.
2. Use a command such as the following to write out the path names in the `fsimage`:

```
$ hdfs oiv -i FSIMAGE -o /tmp/YYYY-MM-DD_FSIMAGE.txt
```

3. Use a command such as the following to find the path names of any symbolic links listed in `/tmp/YYYY-MM-DD_FSIMAGE.txt` and write them out to the file `/tmp/symlinks.txt`:

```
$ grep -- "->" /tmp/YYYY-MM-DD_FSIMAGE.txt > /tmp/symlinks.txt
```

4. Delete any symbolic links listed in `/tmp/symlinks.txt`.

Check Hue Table Sizes and Cleanup if Necessary

When upgrading from CDH 4 to CDH 5, Hue upgrade can take a very long time if the `beeswax_queryhistory`, `beeswax_savedquery`, and `oozie_job` tables are larger than 1000 records. You can reduce the upgrade time by running a script to reduce the size of the Hue database:

1. Stop the Hue service.
2. Back up the Hue database.
3. Download the [history cleanup script](#) to the host running the Hue Server.
4. Run the following as root:

- **parcel installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l  
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"  
/opt/cloudera/parcels/CDH/share/hue/build/env/bin/hue shell
```

- **package installation**

```
export HUE_CONF_DIR="/var/run/cloudera-scm-agent/process/`ls -l  
/var/run/cloudera-scm-agent/process | grep HUE | sort -n | tail -1`"  
/usr/share/hue/build/env/bin/hue shell
```

5. Run the downloaded script in the Hue shell.

Considerations for Secure Clusters

If you are upgrading a cluster that has Kerberos security enabled, you must do the following:

- Before starting the upgrade, make sure your installation is properly configured according to the instructions in the installation and configuration sections of the [Cloudera Security](#) guide.
- Before shutting down Hadoop services, put the NameNode into safe mode and perform a `saveNamespace` operation; see the instructions on [backing up the metadata](#).

High Availability

In CDH 5, you can configure high availability both for the NameNode and the JobTracker or ResourceManager.

- For more information and instructions on setting up a new HA configuration, see [High Availability](#).


Important:

If you configure [HA for the NameNode](#), do not install `hadoop-hdfs-secondarynamenode`. After completing the [HDFS HA software configuration](#), follow the installation instructions in [Deploying HDFS High Availability](#).

- To upgrade an existing configuration, follow the instructions in [Upgrading to CDH 5](#) on page 169.

Upgrading to CDH 5


Note: Are you on the right page?

Use the instructions on this page only to upgrade from CDH 4.

To upgrade from an earlier CDH 5 release to the latest version, use the appropriate instructions:

- [Upgrading from a Release Lower than CDH 5.4.0 to the Latest Release](#) on page 199
- [Upgrading from CDH 5.4.0 or Higher to the Latest Release](#) on page 186


Important:

- To upgrade from CDH 4, you must uninstall CDH 4 and then install CDH 5. Allow sufficient time for this, and do the required backup and preparation as described below.
- If you have configured HDFS HA with NFS shared storage, do not proceed. This configuration is not supported on CDH 5; Quorum-based storage is the only supported HDFS HA configuration on CDH 5. [Unconfigure](#) your NFS shared storage configuration before you attempt to upgrade.


Note: Running Services

Use the `service` command to start, stop, and restart CDH components, instead of running scripts in `/etc/init.d` directly. The `service` command creates a predictable environment by setting the current working directory to `/` and removing most environment variables (passing only `LANG` and `TERM`). With `/etc/init.d`, existing environment variables remain in force and can produce unpredictable results. When you install CDH from packages, `service` is installed as part of the Linux Standard Base (LSB).

Back Up Configuration Data and Stop Services

- Put the NameNode into safe mode and save the `fsimage`:

- Put the NameNode (or active NameNode in an HA configuration) into safe mode:

```
$ sudo -u hdfs hdfs dfsadmin -safemode enter
```

- Perform a `saveNamespace` operation:

```
$ sudo -u hdfs hdfs dfsadmin -saveNamespace
```

This results in a new `fsimage` written with no edit log entries.

- For each component, back up configuration data, databases, and other important files.

3. Shut down the Hadoop services across your entire cluster:

```
for x in `cd /etc/init.d ; ls hadoop-*` ; do sudo service $x stop ; done
```

4. Check each host to make sure that no processes are running as the `hdfs` or `mapred` users from root:

```
# ps -aef | grep java
```

Back Up HDFS Metadata



Important:

Do this step when you are sure that all Hadoop services have been shut down. **To create a consistent backup, verify that the NameNode service is not running.**

To back up the HDFS metadata on the NameNode machine:



Note:

- Cloudera recommends backing up HDFS metadata on a regular basis, as well as before a major upgrade.
- `dfs.name.dir` is deprecated but still works; `dfs.namenode.name.dir` is preferred. This example uses `dfs.name.dir`.

1. Find the location of your `dfs.name.dir` (or `dfs.namenode.name.dir`); for example:

```
$ grep -C1 dfs.name.dir /etc/hadoop/conf/hdfs-site.xml
```

You should see something like this:

```
<property>
<name>dfs.name.dir</name>
<value>/mnt/hadoop/hdfs/name</value>
```

2. Back up the directory. The path inside the `<value>` XML element is the path to your HDFS metadata. If you see a comma-separated list of paths, you do not need back up all of them; they store the same data. Back up the first directory, for example, by using the following commands:

```
$ cd /mnt/hadoop/hdfs/name
# tar -cvf /root/nn_backup_data.tar .
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
./image/
./image/fsimage
```



Warning: If you see a filename containing the word `lock`, the NameNode is probably still running. Repeat the preceding steps, starting by shutting down the Hadoop services.

Uninstall the CDH 4 Version of Hadoop



Warning: Do not proceed before you have backed up the HDFS metadata, and the files and databases for the individual components, as instructed in the previous steps.

To uninstall Hadoop:

Run this command on each host:

On RHEL-compatible systems:

```
$ sudo yum remove bigtop-utils bigtop-jsvc bigtop-tomcat sqoop2-client hue-common solr
```

On SLES systems:

```
$ sudo zypper remove bigtop-utils bigtop-jsvc bigtop-tomcat sqoop2-client hue-common solr
```

On Ubuntu systems:

```
sudo apt-get remove bigtop-utils bigtop-jsvc bigtop-tomcat sqoop2-client hue-common solr
```

Remove CDH 4 Repository Files

Remove all Cloudera CDH 4 repository files. For example, on an RHEL or similar system, remove all files in `/etc/yum.repos.d` that have `cloudera` as part of the name.

**Important:**

- Before removing the files, make sure you have not added any custom entries that you want to preserve. (To preserve custom entries, back up the files before removing them.)
- Make sure you remove Impala and Search repository files, as well as the CDH repository file.

Download the Latest Version of CDH 5

**Note:**

For instructions on how to add a CDH 5 yum repository or build your own CDH 5 yum repository, see [Installing the Latest CDH 5 Release](#).

On RHEL-compatible systems:

1. Download the CDH 5 "1-click Install" package (or RPM).

Click the appropriate RPM and **Save File** to a directory with write access (for example, your home directory).

OS Version	Link to CDH 5 RPM
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS/Oracle 6	RHEL/CentOS/Oracle 6 link
RHEL/CentOS/Oracle 7	RHEL/CentOS/Oracle 7 link

2. Install the RPM for all RHEL versions:

```
$ sudo yum --nogpgcheck localinstall cloudera-cdh-5-0.x86_64.rpm
```

**Important:**

Before proceeding, clean cached packages and headers to ensure that your system repos are up-to-date:

```
sudo yum clean all
```

Optionally add a repository key:

- **For RHEL/CentOS/Oracle 5 systems:**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **For RHEL/CentOS/Oracle 6 systems:**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

On SLES systems:

1. Download the CDH 5 "1-click Install" package.

Download the [RPM file](#), choose **Save File**, and save it to a directory to which you have write access (for example, your home directory).

2. Install the RPM:

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

3. Update your system package index by running the following:

```
$ sudo zypper refresh
```

**Important:**

Before proceeding, clean cached packages and headers to ensure your system repos are up-to-date:

```
sudo zypper clean --all
```

Optionally add a repository key:

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

On Ubuntu and Debian systems:

1. Download the CDH 5 "1-click Install" package:

OS Version	Package Link
Jessie	Jessie package
Wheezy	Wheezy package
Precise	Precise package
Trusty	Trusty package

2. Install the package by doing one of the following:

- Choose **Open with** in the download window to use the package manager.
- Choose **Save File**, save the package to a directory to which you have write access (for example, your home directory), and install it from the command line. For example:

```
sudo dpkg -i cdh5-repository_1.0_all.deb
```



Important: Clean cached packages and headers to ensure that your system repos are up-to-date:

```
sudo apt-get update
```

Optionally add a repository key:

- **For Ubuntu Precise systems:**

```
$ curl -s https://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key  
| sudo apt-key add -
```

- **For Debian Wheezy systems:**

```
$ curl -s https://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key  
| sudo apt-key add -
```

Install CDH 5 with YARN



Note: Skip this step and go to [Install CDH 5 with MRv1](#) on page 174 if you intend to use *only* MRv1.

1. Install and deploy ZooKeeper.



Important: Cloudera recommends that you install (or update) and start a ZooKeeper cluster before proceeding. This is a **requirement** if you are deploying high availability (HA) for the NameNode or JobTracker.

Follow the instructions in [ZooKeeper Installation](#).

2. Install each type of daemon package on the appropriate system, as follows.

Where to install	Install commands
Resource Manager host (analogous to MRv1 JobTracker) running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-yarn-resourcemanager</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-yarn-resourcemanager</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-yarn-resourcemanager</code>
NameNode host(s) running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>

Where to install	Install commands
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>
All cluster hosts except the Resource Manager running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper clean --all; sudo zypper install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
One host in the cluster running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
All client hosts , running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-client</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-client</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-client</code>



Note: The `hadoop-yarn` and `hadoop-hdfs` packages are installed on each system automatically as dependencies of the other packages.

Install CDH 5 with MRv1



Note:

Skip this step if you intend to use *only* YARN. If you are installing both YARN and MRv1, you can skip any packages you have already installed in [Step 6a](#).

To install CDH 5 with MRv1:



Note: If you are also installing YARN, you can skip any packages you have already installed in [Step 6a](#).

1. Install and deploy ZooKeeper.



Important: Cloudera recommends that you install (or update) and start a ZooKeeper cluster before proceeding. This is a **requirement** if you are deploying high availability (HA) for the NameNode or JobTracker.

Follow instructions under [ZooKeeper Installation](#).

2. Install each type of daemon package on the appropriate systems(s), as follows.

Where to install	Install commands
JobTracker host running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-jobtracker</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-jobtracker</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-jobtracker</code>
NameNode host(s) running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>
All cluster hosts except the JobTracker, NameNode, and Secondary (or Standby) NameNode hosts , running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>

Where to install	Install commands
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
All client hosts, running:	
<i>RHEL/CentOS compatible</i>	<code>sudo yum clean all; sudo yum install hadoop-client</code>
<i>SLES</i>	<code>sudo zypper clean --all; sudo zypper install hadoop-client</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-client</code>

Copy the CDH 5 Logging File

Copy over the `log4j.properties` file to your custom directory on each host in the cluster; for example:

```
$ cp /etc/hadoop/conf.empty/log4j.properties /etc/hadoop/conf.my_cluster/log4j.properties
```

In an HA Deployment, Upgrade and Start the JournalNodes

1. Install the JournalNode daemons on each of the machines where they will run.

To install JournalNode on RHEL-compatible systems:

```
$ sudo yum install hadoop-hdfs-journalnode
```

To install JournalNode on Ubuntu and Debian systems:

```
$ sudo apt-get install hadoop-hdfs-journalnode
```

To install JournalNode on SLES systems:

```
$ sudo zypper install hadoop-hdfs-journalnode
```

2. Start the JournalNode daemons on each of the machines where they will run:

```
sudo service hadoop-hdfs-journalnode start
```

Wait for the daemons to start before proceeding to the next step.



Important: The JournalNodes must be up and running CDH 5 before you proceed.

Upgrade the HDFS Metadata

**Note:**

This step differs depending on whether you are upgrading an HDFS HA deployment using Quorum-based storage, or a non-HA deployment using a secondary NameNode. (If you have an HDFS HA deployment using NFS storage, do not proceed; you cannot upgrade that configuration to CDH 5. [Unconfigure](#) your NFS shared storage configuration before you attempt to upgrade.)

- For an HA deployment, do sub-steps 1, 2, and 3 below.
- For a non-HA deployment, do sub-steps 1, 3, and 4 below.

1. To upgrade the HDFS metadata, run the following command on the NameNode. If HA is enabled, do this on the *active NameNode only*, and make sure the JournalNodes have been upgraded to CDH 5 and are up and running before you run the command.

```
$ sudo service hadoop-hdfs-namenode upgrade
```



Important: In an HDFS HA deployment, you do this on *only one* NameNode.

You can watch the progress of the upgrade by running:

```
$ sudo tail -f /var/log/hadoop-hdfs/hadoop-hdfs-namenode-<hostname>.log
```

Look for a line that confirms the upgrade is complete, such as:

```
/var/lib/hadoop-hdfs/cache/hadoop/dfs/<name> is complete
```



Note: The NameNode upgrade process can take a while depending on how many files you have.

- 2.



Note: Do this step only in a non-HA configuration. Otherwise, skip to step 3.

step 3.

Wait for NameNode to exit safe mode, and then restart the standby NameNode.

- If Kerberos is enabled:

```
$ kinit -kt /path/to/hdfs.keytab hdfs/<fully.qualified.domain.name@YOUR-REALM.COM> &&
hdfs namenode -bootstrapStandby
```

```
$ sudo service hadoop-hdfs-namenode start
```

- If Kerberos is not enabled:


```
$ sudo -u hdfs hdfs namenode -bootstrapStandby
$ sudo service hadoop-hdfs-namenode start
```

For more information about the `haadmin -failover` command, see [Administering an HDFS High Availability Cluster](#).

3. Start up each DataNode

:

```
$ sudo service hadoop-hdfs-datanode start
```

4.  **Note:** Do this step only in a non-HA configuration. Otherwise, skip to [Start YARN or MapReduce MRv1](#) on page 178.

Wait for the NameNode to exit safe mode, and then start the secondary NameNode.

- a. To check that the NameNode has exited safe mode, look for messages in the log file or the NameNode web interface that read "...no longer in safe mode."
- b. To start the secondary NameNode (if used), enter the following command on the secondary NameNode host:

```
$ sudo service hadoop-hdfs-secondarynamenode start
```


- c. To complete the cluster upgrade, follow the steps below.

Start YARN or MapReduce MRv1

You are now ready to start and test MRv1 or YARN.

For YARN	or For MRv1
Start YARN and the MapReduce JobHistory Server	Start MRv1
Verify basic cluster operation	Verify basic cluster operation

Start MapReduce with YARN

 **Important:** Make sure you are not trying to run MRv1 and YARN on the same set of hosts at the same time. This degrades performance and may result in an unstable MapReduce cluster deployment.

After you have verified that HDFS is operating correctly, you can start YARN. First, create directories and set the correct permissions.


 **Note:** For more information see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

Create a history directory and set permissions; for example:

```
sudo -u hdfs hadoop fs -mkdir /user/history
sudo -u hdfs hadoop fs -chmod -R 1777 /user/history
sudo -u hdfs hadoop fs -chown yarn /user/history
```

Create the `/var/log/hadoop-yarn` directory and set ownership:

```
$ sudo -u hdfs hadoop fs -mkdir /var/log/hadoop-yarn
$ sudo -u hdfs hadoop fs -chown yarn:mapred /var/log/hadoop-yarn
```

 **Note:** You need to create this directory because it is the parent of `/var/log/hadoop-yarn/apps` which is explicitly configured in the `yarn-site.xml`.

Verify the directory structure, ownership, and permissions:

```
$ sudo -u hdfs hadoop fs -ls -R /
```

You should see:

```
drwxrwxrwt - hdfs supergroup 0 2012-04-19 14:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2012-05-31 10:26 /user
drwxrwxrwt - yarn supergroup 0 2012-04-19 14:31 /user/history
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var/log
drwxr-xr-x - yarn mapred 0 2012-05-31 15:31 /var/log/hadoop-yarn
```

To start YARN, start the ResourceManager and NodeManager services:



Note: Make sure you always start ResourceManager before starting NodeManager services.

On the ResourceManager system:

```
$ sudo service hadoop-yarn-resourcemanager start
```

On each NodeManager system (typically the same ones where DataNode service runs):

```
$ sudo service hadoop-yarn-nodemanager start
```

To start the MapReduce JobHistory Server

On the MapReduce JobHistory Server system:

```
$ sudo service hadoop-mapreduce-historyserver start
```

For each user who will submit MapReduce jobs using MapReduce v2 (YARN), or running Pig, Hive, or Sqoop in a YARN installation, make sure that the `HADOOP_MAPRED_HOME` environment variable is set correctly as follows:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

Verify basic cluster operation for YARN.

At this point, your cluster is upgraded and ready to run jobs. Before running production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.



Note:

For important configuration information, see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

1. Create a home directory on HDFS for the user running the job (for example, joe):

```
$ sudo -u hdfs hadoop fs -mkdir /user/joe
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Do the following steps as user joe.

2. Make a directory in HDFS called `input` and copy some XML files to it by running the following commands in pseudo-distributed mode:

```
$ hadoop fs -mkdir input
$ hadoop fs -put /etc/hadoop/conf/*.xml input
$ hadoop fs -ls input
```

Upgrading Unmanaged CDH Using the Command Line

```
Found 3 items:
-rw-r--r--  1 joe supergroup      1348 2012-02-13 12:21 input/core-site.xml
-rw-r--r--  1 joe supergroup      1913 2012-02-13 12:21 input/hdfs-site.xml
-rw-r--r--  1 joe supergroup      1001 2012-02-13 12:21 input/mapred-site.xml
```

3. Set HADOOP_MAPRED_HOME for user joe:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

4. Run an example Hadoop job to grep with a regular expression in your input data.

```
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar grep input output23 'dfs[a-z.]+'
```

5. After the job completes, find the output in the HDFS directory named output23 because you specified that output directory to Hadoop.

```
$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - joe supergroup  0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x  - joe supergroup  0 2009-08-18 18:38 /user/joe/output23
```

There is a new directory called output23.

6. List the output files.

```
$ hadoop fs -ls output23
Found 2 items
drwxr-xr-x  - joe supergroup      0 2009-02-25 10:33 /user/joe/output23/_SUCCESS
-rw-r--r--  1 joe supergroup    1068 2009-02-25 10:33 /user/joe/output23/part-r-00000
```

7. Read the results in the output file.

```
$ hadoop fs -cat output23/part-r-00000 | head
1 dfs.safemode.min.datanodes
1 dfs.safemode.extension
1 dfs.replication
1 dfs.permissions.enabled
1 dfs.namenode.name.dir
1 dfs.namenode.checkpoint.dir
1 dfs.datanode.data.dir
```

You have confirmed your cluster is successfully running CDH 5.



Important: If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients as well.

Start MapReduce (MRv1)



Important: Make sure you are not running MRv1 and YARN on the same set of hosts at the same time. This degrades performance and may result in an unstable MapReduce cluster deployment.

After you have verified that HDFS is operating correctly, you can start MapReduce. On each TaskTracker system:

```
$ sudo service hadoop-0.20-mapreduce-tasktracker start
```

On the JobTracker system:

```
$ sudo service hadoop-0.20-mapreduce-jobtracker start
```

Verify that the JobTracker and TaskTracker started properly.

```
$ sudo jps | grep Tracker
```

If the permissions of directories are not configured correctly, the JobTracker and TaskTracker processes start and immediately fail. If this happens, check the JobTracker and TaskTracker logs and set the permissions correctly.



Important: For each user submitting MapReduce jobs using MapReduce v1 (MRv1), or running Pig, Hive, or Sqoop in an MRv1 installation, set the `HADOOP_MAPRED_HOME` environment variable as follows:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-0.20-mapreduce
```

Verify basic cluster operation for MRv1

Your cluster is upgraded and ready to run jobs. Before running production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.

1. Create a home directory on HDFS for the user running the job (for example, `joe`):

```
$ sudo -u hdfs hadoop fs -mkdir /user/joe
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Do the following steps as user `joe`.

2. Make a directory in HDFS called `input` and copy some XML files into it by running the following commands:

```
$ hadoop fs -mkdir input
$ hadoop fs -put /etc/hadoop/conf/*.xml input
$ hadoop fs -ls input
Found 3 items:
-rw-r--r--  1 joe supergroup    1348 2012-02-13 12:21 input/core-site.xml
-rw-r--r--  1 joe supergroup    1913 2012-02-13 12:21 input/hdfs-site.xml
-rw-r--r--  1 joe supergroup    1001 2012-02-13 12:21 input/mapred-site.xml
```

3. Set `HADOOP_MAPRED_HOME` for user `joe`:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-0.20-mapreduce/
```

4. Run an example Hadoop job to grep with a regular expression in your input data.

```
$ /usr/bin/hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples.jar grep input
output 'dfs[a-z.]+'
```

5. After the job completes, find the output in the HDFS directory named `output` because you specified that output directory to Hadoop.

```
$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - joe supergroup  0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x  - joe supergroup  0 2009-08-18 18:38 /user/joe/output
```

ere is a new directory called `output`.

6. List the output files.

```
$ hadoop fs -ls output
Found 2 items
drwxr-xr-x -   joe supergroup      0 2009-02-25 10:33 /user/joe/output/_logs
-rw-r--r-- 1   joe supergroup 1068 2009-02-25 10:33 /user/joe/output/part-00000
-rw-r--r-- 1   joe supergroup      0 2009-02-25 10:33 /user/joe/output/_SUCCESS
```

7. Read the results in the output file; for example:

```
$ hadoop fs -cat output/part-00000 | head
1      dfs.datanode.data.dir
1      dfs.namenode.checkpoint.dir
1      dfs.namenode.name.dir
1      dfs.replication
1      dfs.safemode.extension
1      dfs.safemode.min.datanodes
```

You have confirmed your cluster is successfully running CDH 5.



Important:

If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients as well.

Set the Sticky Bit

For security reasons, Cloudera strongly recommends you set the sticky bit on directories if you have not already done so.

The sticky bit prevents anyone except the superuser, directory owner, or file owner from deleting or moving the files in a directory. (Setting the sticky bit for a file has no effect.) Do this for directories such as `/tmp`. (For instructions on creating `/tmp` and setting its permissions, see [Create the /tmp Directory](#) .

Reinstall CDH 5 Components

When upgrading CDH, Cloudera strongly recommends that you upgrade your client jars to match. For help on finding matching artifacts, see [Using the CDH 5 Maven Repository](#).

CDH 5 Components

Use the following sections to install or upgrade CDH 5 components:

- [Crunch Installation](#)
- [Flume Installation](#)
- [HBase Installation](#)
- [HCatalog Installation](#)
- [Hive Installation](#)
- [HttpFS Installation](#)
- [Hue Installation](#)
- [Impala Installation](#)
- [KMS Installation and Upgrade](#)
- [Mahout Installation](#)
- [Oozie Installation](#)
- [Pig Installation](#)
- [Search Installation](#)
- [Sentry Installation](#)
- [Snappy Installation](#)
- [Spark Installation](#)

- [Sqoop 1 Installation](#)
- [Sqoop 2 Installation](#)
- [Whirr Installation](#)
- [ZooKeeper Installation](#)

See also the instructions for [installing or updating LZO](#).

Apply Configuration File Changes



Important:

During uninstall, the package manager renames any configuration files you have modified from `<file>` to `<file>.rpmsave`. During reinstall, the package manager creates a new `<file>` with applicable defaults. You must apply any changes captured in the original CDH 4 configuration file to the new CDH 5 configuration file. In Ubuntu and Debian upgrades, a file is not installed if a version of that file already exists on the system, and you are prompted to resolve conflicts; for details, see [Automatic handling of configuration files by dpkg](#).

For example, if you have modified your CDH 4 `zoo.cfg` configuration file (`/etc/zookeeper.dist/zoo.cfg`), RPM uninstall and reinstall (using `yum remove`) renames and preserves a copy of your modified `zoo.cfg` as `/etc/zookeeper.dist/zoo.cfg.rpmsave`. You should compare this to the new `/etc/zookeeper/conf/zoo.cfg` and resolve any differences that should be carried forward (typically where you have changed property value defaults). Do this for each component you upgrade to CDH 5.

Finalize the HDFS Metadata Upgrade

To finalize the HDFS metadata upgrade you began earlier in this procedure, proceed as follows:

1. Make sure you are satisfied that the CDH 5 upgrade has succeeded and everything is running smoothly. To determine if you can finalize, run important workloads and ensure that they are successful.



Warning: Do not proceed until you are sure you are satisfied with the new deployment. Once you have finalized the HDFS metadata, you cannot revert to an earlier version of HDFS.



Note:

- If you need to restart the NameNode during this period (after having begun the upgrade process, but before you have run `finalizeUpgrade`), restart your NameNode without the `-upgrade` option.
- Verifying that you are ready to finalize the upgrade can take a long time. Make sure you have enough free disk space, keeping in mind that the following behavior continues until the upgrade is finalized:
 - Deleting files does not free up disk space.
 - Using the balancer causes all moved replicas to be duplicated.
 - All on-disk data representing the NameNodes metadata is retained, which could more than double the amount of space required on the NameNode and JournalNode disks.

2. Finalize the HDFS metadata upgrade using one of the following commands, depending on whether Kerberos is enabled (see [Configuring Hadoop Security in CDH 5](#)).



Important: In an HDFS HA deployment, make sure that both the NameNodes and all of the JournalNodes are up and functioning normally before you proceed.

Upgrading Unmanaged CDH Using the Command Line

- If Kerberos is enabled:

```
$ kinit -kt /path/to/hdfs.keytab hdfs/<fully.qualified.domain.name@YOUR-REALM.COM> &&  
hdfs dfsadmin -finalizeUpgrade
```

- If Kerberos is not enabled:

```
$ sudo -u hdfs hdfs dfsadmin -finalizeUpgrade
```



Note: After the metadata upgrade completes, the `previous/` and `blocksBeingWritten/` directories in the DataNode data directories are not cleared until the DataNodes are restarted.

Upgrading from an Earlier CDH 5 Release to the Latest Release



Important:

- If you are using Cloudera Manager to manage CDH, *do not* use the instructions in this section. Follow the directions in [Upgrading CDH and Managed Services Using Cloudera Manager](#) on page 54 to upgrade to the latest version of CDH 5 in a Cloudera Manager deployment.
- The instructions in this section describe how to upgrade to the latest CDH 5 release from an earlier CDH 5 release. If you are upgrading from a CDH 4 release, use the instructions in [Upgrading from CDH 4 to CDH 5](#) on page 167 instead.
- MRv1 and YARN share a common set of configuration files, so it is safe to *configure* both of them. Cloudera does not recommend running MapReduce MRv1 and YARN daemons on the same hosts at the same time. If you want to easily switch between MapReduce MRv1 and YARN, use Cloudera Manager [to manage these services](#).



Note: Running Services

Use the `service` command to start, stop, and restart CDH components, instead of running scripts in `/etc/init.d` directly. The `service` command creates a predictable environment by setting the current working directory to `/` and removing most environment variables (passing only `LANG` and `TERM`). With `/etc/init.d`, existing environment variables remain in force and can produce unpredictable results. When you install CDH from packages, `service` is installed as part of the Linux Standard Base (LSB).

Important Tasks

- Upgrading from any release earlier than CDH 5.4.0 to CDH 5.4.0 or later requires an HDFS metadata upgrade.
- Upgrading from a release earlier than 5.2.0 requires all of the following:
 - Upgrade HDFS metadata
 - Upgrade the Sentry database
 - Upgrade the Hive database
 - Upgrade the Sqoop 2 database

Make sure you also do the following tasks that are required for every upgrade:

- Upgrade the Oozie database and shared library.
- If you have uploaded the Spark assembly JAR file to HDFS, upload the new version of the file.

Each of these tasks is described in context as you proceed through the upgrade. The following sections provide information and instructions:

- [Before Upgrading to the Latest Release of CDH](#) on page 185
- [Upgrading from CDH 5.4.0 or Higher to the Latest Release](#) on page 186
- [Upgrading from a Release Lower than CDH 5.4.0 to the Latest Release](#) on page 199

Before Upgrading to the Latest Release of CDH



Note:

- Before upgrading, read about the latest [Incompatible Changes](#) and [Known Issues and Workarounds in CDH 5](#) in the [CDH 5 Release Notes](#).



Warning: Make sure to read [Install and Upgrade Known Issues](#).

- If you are upgrading a cluster that is part of a production system, plan ahead. For production clusters, Cloudera recommends allocating up to a full day maintenance window to perform the upgrade, depending on the number of hosts, the amount of experience you have with Hadoop and Linux, and the particular hardware you are using.
- The instructions in this section assume you are upgrading a multi-node cluster. If you are running a pseudo-distributed (single-machine) cluster, Cloudera recommends that you copy your data from the cluster, remove the old CDH release, install Hadoop from CDH 5, and then restore your data.
- If you have a multi-node cluster running an earlier version of CDH 5, use the appropriate instructions to upgrade your cluster to the latest version:
 - [Upgrading from CDH 5.4.0 or Higher to the Latest Release](#) on page 186
 - [Upgrading from a Release Lower than CDH 5.4.0 to the Latest Release](#) on page 199

Troubleshooting: Upgrading hadoop-kms from 5.2.x and 5.3.x Releases on SLES

This section describes issues that affect SLES upgrades from 5.2.x releases earlier than 5.2.4, and from 5.3.x releases earlier than 5.3.2.

Problem

The problem occurs when you try to upgrade the `hadoop-kms` package, for example:

```
Installing: hadoop-kms-2.5.0+cdh5.3.2+801-1.cdh5.3.2.p0.224.sles11 [error]
12:54:19 Installation of hadoop-kms-2.5.0+cdh5.3.2+801-1.cdh5.3.2.p0.224.sles11 failed:
12:54:19 (with --nodeps --force) Error: Subprocess failed. Error: RPM failed: warning:
/var/cache/zypp/packages/cdh/RPMS/x86_64/hadoop-kms-2.5.0+cdh5.3.2+801-1.cdh5.3.2.p0.224.sles11.x86_64.rpm:
Header V4 DSA signature: NOKEY, key ID e8f86acd
12:54:19 error: %postun(hadoop-kms-2.5.0+cdh5.3.1+791-1.cdh5.3.1.p0.17.sles11.x86_64)
scriptlet failed, exit status 1
12:54:19
```



Note:

- The `hadoop-kms` package is not installed automatically with CDH, so you encounter this error only if you are explicitly upgrading an existing version of KMS.
- The examples in this section show an upgrade from CDH 5.3.x; the 5.2.x case looks very similar.

Upgrading Unmanaged CDH Using the Command Line

What to Do

If you see an error similar to the one in the example above, proceed as follows:

1. Abort or ignore the error (either option works):

```
Abort, retry, ignore? [a/r/i] (a): i
```

2. Perform cleanup:

a. `# rpm -qa hadoop-kms`

You will see two versions of `hadoop-kms`; for example:

```
hadoop-kms-2.5.0+cdh5.3.1+791-1.cdh5.3.1.p0.17.sles11
hadoop-kms-2.5.0+cdh5.3.2+801-1.cdh5.3.2.p0.224.sles11
```

- b. Remove the older version, in this example

`hadoop-kms-2.5.0+cdh5.3.1+791-1.cdh5.3.1.p0.17.sles11:`

```
# rpm -e --noscripts hadoop-kms-2.5.0+cdh5.3.1+791-1.cdh5.3.1.p0.17.sles11
```

3. Verify that the older version of the package has been removed:

```
# rpm -qa hadoop-kms
```

You should now see only the newer package:

```
hadoop-kms-2.5.0+cdh5.3.2+801-1.cdh5.3.2.p0.224.sles11
```

Upgrading from CDH 5.4.0 or Higher to the Latest Release

Use the instructions that follow to upgrade from CDH 5.4.0 or higher to the latest version of CDH 5.



Important: Are you on the right page?

Use the instructions on this page *only* to upgrade from CDH 5.4.0 or higher. Upgrades from a release earlier than CDH 5.4.0 require an HDFS metadata upgrade.

If you are *not* currently running CDH 5.4.0 or higher:

- Use [these instructions](#) to upgrade from a CDH 5 release earlier than CDH 5.4.0;
- Use [these instructions](#) to upgrade from a CDH 4 release.

Step 1: Prepare the cluster for the upgrade

1. Put the NameNode into safe mode and save the `fsimage`:

- a. Put the NameNode (or active NameNode in an HA configuration) into safe mode:

```
$ sudo -u hdfs hdfs dfsadmin -safemode enter
```

- b. Perform a `saveNamespace` operation:

```
$ sudo -u hdfs hdfs dfsadmin -saveNamespace
```

This will result in a new `fsimage` being written out with no edit log entries.

- c. With the NameNode still in safe mode, shut down all services as instructed below.

2. Shut down Hadoop services across your entire cluster by running the following command on every host in your cluster:

```
$ for x in `cd /etc/init.d ; ls hadoop-*` ; do sudo service $x stop ; done
```

3. Check each host to make sure that there are no processes running as the `hdfs`, `yarn`, `mapred` or `httpfs` users from root:

```
# ps -aef | grep java
```


Important:

When you are sure that all Hadoop services have been shut down, do the following step. **It is particularly important that the NameNode service is not running so that you can make a consistent backup.**

4. Back up the HDFS metadata on the NameNode machine, as follows.


Note:

- Cloudera recommends backing up HDFS metadata on a regular basis, as well as before a major upgrade.
- `dfs.name.dir` is deprecated but still works; `dfs.namenode.name.dir` is preferred. This example uses `dfs.name.dir`.

- a. Find the location of your `dfs.name.dir` (or `dfs.namenode.name.dir`); for example:

```
$ grep -C1 dfs.name.dir /etc/hadoop/conf/hdfs-site.xml
<property> <name>dfs.name.dir</name> <value>/mnt/hadoop/hdfs/name</value>
</property>
```

- b. Back up the directory. The path inside the `<value>` XML element is the path to your HDFS metadata. If you see a comma-separated list of paths, there is no need to back up all of them; they store the same data. Back up the first directory, for example, by using the following commands:

```
$ cd /mnt/hadoop/hdfs/name
# tar -cvf /root/nn_backup_data.tar .
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
./image/
./image/fsimage
```


Important:

If you see a file containing the word *lock*, the NameNode is probably still running. Repeat the preceding steps from the beginning; start at Step 1 and shut down the Hadoop services.

Step 2: If necessary, download the CDH 5 "1-click" package on each host in your cluster

Before you begin: Check whether you have the CDH 5 "1-click" repository installed.

Upgrading Unmanaged CDH Using the Command Line

- On Red Hat/CentOS-compatible and SLES systems:

```
rpm -q cdh5-repository
```

If you are upgrading from CDH 5 Beta 1 or higher, you should see:

```
cdh5-repository-1-0
```

In this case, skip to [Step 3](#). If instead you see:

```
package cdh5-repository is not installed
```

proceed with [this step](#).

- On Ubuntu and Debian systems:

```
dpkg -l | grep cdh5-repository
```

If the repository is installed, skip to [Step 3](#); otherwise proceed with [this step](#).

If the CDH 5 "1-click" repository is not already installed on each host in the cluster, follow the instructions below for that host's operating system:

[Instructions for Red Hat-compatible systems](#)

[Instructions for SLES systems](#)

[Instructions for Ubuntu and Debian systems](#)

On Red Hat-compatible systems:

1. Download the CDH 5 "1-click Install" package (or RPM).

Click the appropriate RPM and **Save File** to a directory with write access (for example, your home directory).

OS Version	Link to CDH 5 RPM
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS/Oracle 6	RHEL/CentOS/Oracle 6 link
RHEL/CentOS/Oracle 7	RHEL/CentOS/Oracle 7 link

2. Install the RPM for all RHEL versions:

```
$ sudo yum --nogpgcheck localinstall cloudera-cdh-5-0.x86_64.rpm
```



Note:

For instructions on how to add a CDH 5 yum repository or build your own CDH 5 yum repository, see [Installing CDH 5 On Red Hat-compatible systems](#).



Note: Clean repository cache.

Before proceeding, clean cached packages and headers to ensure your system repos are up-to-date:

```
sudo yum clean all
```

On SLES systems:

1. Download the CDH 5 "1-click Install" package.

Download the [RPM file](#), choose **Save File**, and save it to a directory to which you have write access (for example, your home directory).

2. Install the RPM:

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

3. Update your system package index by running the following:

```
$ sudo zypper refresh
```

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```



Note:

For instructions on how to add a repository or build your own repository, see [Installing CDH 5 on SLES Systems](#).



Note: Clean repository cache.

Before proceeding, clean cached packages and headers to ensure your system repos are up-to-date:

```
sudo zypper clean --all
```

On Ubuntu and Debian systems:

1. Download the CDH 5 "1-click Install" package:

OS Version	Package Link
Jessie	Jessie package
Wheezy	Wheezy package
Precise	Precise package
Trusty	Trusty package

2. Install the package by doing one of the following:

- Choose **Open with** in the download window to use the package manager.
- Choose **Save File**, save the package to a directory to which you have write access (for example, your home directory), and install it from the command line. For example:

```
sudo dpkg -i cdh5-repository_1.0_all.deb
```



Note:

For instructions on how to add a repository or build your own repository, see [Installing CDH 5 on Ubuntu Systems](#).



Important: Clean cached packages and headers to ensure that your system repos are up-to-date:
`sudo apt-get update`

Step 3: Upgrade the Packages on the Appropriate Hosts

Upgrade [MRv1](#), [YARN](#), or both, depending on what you intend to use.



Note:

- Remember that you can install and configure both MRv1 and YARN, but you should not run them both on the same set of hosts at the same time.
- If you are using [HA for the NameNode](#), do not install `hadoop-hdfs-secondarynamenode`.

Before installing MRv1 or YARN: (Optionally) add a repository key on each system in the cluster, if you have not already done so. Add the Cloudera Public GPG Key to your repository by executing one of the following commands:

- For Red Hat/CentOS/Oracle 5 systems:**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- For Red Hat/CentOS/Oracle 6 systems:**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- For all SLES systems:**

```
$ sudo rpm --import  
https://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- For Ubuntu Precise systems:**

```
$ curl -s  
https://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key  
| sudo apt-key add -
```

- For Debian Wheezy systems:**

```
$ curl -s  
https://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key  
| sudo apt-key add -
```

Step 3a: If you are using MRv1, upgrade the MRv1 packages on the appropriate hosts.

Skip this step if you are using [YARN](#) exclusively. Otherwise upgrade each type of daemon package on the appropriate hosts as follows:

1. Install and deploy ZooKeeper:



Important:

Cloudera recommends that you install (or update) and start a ZooKeeper cluster before proceeding. This is a **requirement** if you are deploying high availability (HA) for the NameNode or JobTracker.

Follow instructions under [ZooKeeper Installation](#).

2. Install each type of daemon package on the appropriate systems(s), as follows.

Where to install	Install commands
JobTracker host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-jobtracker</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-jobtracker</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-jobtracker</code>
NameNode host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>
All cluster hosts except the JobTracker, NameNode, and Secondary (or Standby) NameNode hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
All client hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-client</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-client</code>

Where to install	Install commands
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-client</code>

Step 3b: If you are using YARN, upgrade the YARN packages on the appropriate hosts.

Skip this step if you are using [MRv1](#) exclusively. Otherwise upgrade each type of daemon package on the appropriate hosts as follows:

1. Install and deploy ZooKeeper:



Important:

Cloudera recommends that you install (or update) and start a ZooKeeper cluster before proceeding. This is a **requirement** if you are deploying high availability (HA) for the NameNode or JobTracker.

Follow instructions under [ZooKeeper Installation](#).

2. Install each type of daemon package on the appropriate systems(s), as follows.

Where to install	Install commands
Resource Manager host (analogous to MRv1 JobTracker) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-yarn-resourcemanager</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-yarn-resourcemanager</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-yarn-resourcemanager</code>
NameNode host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>

Where to install	Install commands
All cluster hosts except the Resource Manager (analogous to MRv1 TaskTrackers) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
One host in the cluster running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
All client hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-client</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-client</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-client</code>


Note:

The `hadoop-yarn` and `hadoop-hdfs` packages are installed on each system automatically as dependencies of the other packages.

Step 4: In an HA Deployment, Upgrade and Start the JournalNodes

1. Install the JournalNode daemons on each of the machines where they will run.

To install JournalNode on RHEL-compatible systems:

```
$ sudo yum install hadoop-hdfs-journalnode
```

To install JournalNode on Ubuntu and Debian systems:

```
$ sudo apt-get install hadoop-hdfs-journalnode
```


To install JournalNode on SLES systems:

```
$ sudo zypper install hadoop-hdfs-journalnode
```

2. Start the JournalNode daemons on each of the machines where they will run:

```
sudo service hadoop-hdfs-journalnode start
```

Wait for the daemons to start before proceeding to the next step.

**Important:**
The JournalNodes must be up and running CDH 5 before you proceed.

Step 5: Start HDFS


```
for x in `cd /etc/init.d ; ls hadoop-hdfs-*` ; do sudo service $x start ; done
```

Step 6: Start MapReduce (MRv1) or YARN

You are now ready to start and test MRv1 or YARN.

For MRv1	For YARN
Start MRv1	Start YARN and the MapReduce JobHistory Server
Verify basic cluster operation	Verify basic cluster operation

Step 6a: Start MapReduce (MRv1)

**Important:**
Make sure you are not trying to run MRv1 and YARN on the same set of hosts at the same time. This is not recommended; it will degrade performance and may result in an unstable MapReduce cluster deployment. Steps 6a and 6b are mutually exclusive.

After you have verified HDFS is operating correctly, you are ready to start MapReduce. On each TaskTracker system:

```
$ sudo service hadoop-0.20-mapreduce-tasktracker start
```

On the JobTracker system:

```
$ sudo service hadoop-0.20-mapreduce-jobtracker start
```

Verify that the JobTracker and TaskTracker started properly.

```
$ sudo jps | grep Tracker
```

If the permissions of directories are not configured correctly, the JobTracker and TaskTracker processes start and immediately fail. If this happens, check the JobTracker and TaskTracker logs and set the permissions correctly.

Verify basic cluster operation for MRv1

At this point your cluster is upgraded and ready to run jobs. Before running your production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.


Note:

For important configuration information, see [Deploying MapReduce v1 \(MRv1\) on a Cluster](#).

1. Create a home directory on HDFS for the user who will be running the job (for example, joe):

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/joe
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Do the following steps as the user joe.

2. Make a directory in HDFS called `input` and copy some XML files into it by running the following commands:

```
$ hadoop fs -mkdir input
$ hadoop fs -put /etc/hadoop/conf/*.xml input
$ hadoop fs -ls input
Found 3 items:
-rw-r--r-- 1 joe supergroup 1348 2012-02-13 12:21 input/core-site.xml
-rw-r--r-- 1 joe supergroup 1913 2012-02-13 12:21 input/hdfs-site.xml
-rw-r--r-- 1 joe supergroup 1001 2012-02-13 12:21 input/mapred-site.xml
```

3. Run an example Hadoop job to grep with a regular expression in your input data.

```
$ /usr/bin/hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples.jar grep input
output 'dfs[a-z.]+'
```

4. After the job completes, you can find the output in the HDFS directory named `output` because you specified that output directory to Hadoop.

```
$ hadoop fs -ls
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x - joe supergroup 0 2009-08-18 18:38 /user/joe/output
```

You can see that there is a new directory called `output`.

5. List the output files.

```
$ hadoop fs -ls output
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-02-25 10:33 /user/joe/output/_logs
-rw-r--r-- 1 joe supergroup 1068 2009-02-25 10:33 /user/joe/output/part-00000
-rw-r--r-- 1 joe supergroup 0 2009-02-25 10:33 /user/joe/output/_SUCCESS
```

6. Read the results in the output file; for example:

```
$ hadoop fs -cat output/part-00000 | head
1 dfs.datanode.data.dir
1 dfs.namenode.checkpoint.dir
1 dfs.namenode.name.dir
1 dfs.replication
1 dfs.safemode.extension
1 dfs.safemode.min.datanodes
```

You have now confirmed your cluster is successfully running CDH 5.

**Important:**

If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients as well.

Step 6b: Start MapReduce with YARN

**Important:**

Make sure you are not trying to run MRv1 and YARN on the same set of hosts at the same time. This is not recommended, especially in a cluster that is not managed by Cloudera Manager; it will degrade your performance and may result in an unstable MapReduce cluster deployment. Steps 6a and 6b are mutually exclusive.

After you have verified HDFS is operating correctly, you are ready to start YARN. First, if you have not already done so, create directories and set the correct permissions.



Note: For more information see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

Create a history directory and set permissions; for example:

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/history
$ sudo -u hdfs hadoop fs -chmod -R 1777 /user/history
$ sudo -u hdfs hadoop fs -chown yarn /user/history
```

Create the `/var/log/hadoop-yarn` directory and set ownership:

```
$ sudo -u hdfs hadoop fs -mkdir -p /var/log/hadoop-yarn
$ sudo -u hdfs hadoop fs -chown yarn:mapred /var/log/hadoop-yarn
```



Note: You need to create this directory because it is the parent of `/var/log/hadoop-yarn/apps` which is explicitly configured in the `yarn-site.xml`.

Verify the directory structure, ownership, and permissions:

```
$ sudo -u hdfs hadoop fs -ls -R /
```

You should see:

```
drwxrwxrwt - hdfs supergroup 0 2012-04-19 14:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2012-05-31 10:26 /user
drwxrwxrwt - yarn supergroup 0 2012-04-19 14:31 /user/history
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var/log
drwxr-xr-x - yarn mapred 0 2012-05-31 15:31 /var/log/hadoop-yarn
```

To start YARN, start the ResourceManager and NodeManager services:

**Note:**

Make sure you always start ResourceManager before starting NodeManager services.

On the ResourceManager system:

```
$ sudo service hadoop-yarn-resourcemanager start
```

On each NodeManager system (typically the same ones where DataNode service runs):

```
$ sudo service hadoop-yarn-nodemanager start
```

To start the MapReduce JobHistory Server

On the MapReduce JobHistory Server system:

```
$ sudo service hadoop-mapreduce-historyserver start
```

For each user who will be submitting MapReduce jobs using MapReduce v2 (YARN), or running Pig, Hive, or Sqoop 1 in a YARN installation, set the HADOOP_MAPRED_HOME environment variable as follows:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

Verify basic cluster operation for YARN.

At this point your cluster is upgraded and ready to run jobs. Before running your production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.



Note:

For important configuration information, see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

1. Create a home directory on HDFS for the user who will be running the job (for example, joe):

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/joe
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Do the following steps as the user joe.

2. Make a directory in HDFS called `input` and copy some XML files into it by running the following commands in pseudo-distributed mode:

```
$ hadoop fs -mkdir input
$ hadoop fs -put /etc/hadoop/conf/*.xml input
$ hadoop fs -ls input
Found 3 items:
-rw-r--r-- 1 joe supergroup 1348 2012-02-13 12:21 input/core-site.xml
-rw-r--r-- 1 joe supergroup 1913 2012-02-13 12:21 input/hdfs-site.xml
-rw-r--r-- 1 joe supergroup 1001 2012-02-13 12:21 input/mapred-site.xml
```

3. Set HADOOP_MAPRED_HOME for user joe:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

4. Run an example Hadoop job to `grep` with a regular expression in your input data.

```
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar grep input output23 'dfs[a-z.]+'
```

5. After the job completes, you can find the output in the HDFS directory named `output23` because you specified that output directory to Hadoop.

```
$ hadoop fs -ls
Found 2 items
```

Upgrading Unmanaged CDH Using the Command Line

```
drwxr-xr-x - joe supergroup 0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x - joe supergroup 0 2009-08-18 18:38 /user/joe/output23
```

You can see that there is a new directory called `output23`.

6. List the output files:

```
$ hadoop fs -ls output23
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-02-25 10:33 /user/joe/output23/_SUCCESS
-rw-r--r-- 1 joe supergroup 1068 2009-02-25 10:33 /user/joe/output23/part-r-00000
```

7. Read the results in the output file:

```
$ hadoop fs -cat output23/part-r-00000 | head
1 dfs.safemode.min.datanodes
1 dfs.safemode.extension
1 dfs.replication
1 dfs.permissions.enabled
1 dfs.namenode.name.dir
1 dfs.namenode.checkpoint.dir
1 dfs.datanode.data.dir
```

You have now confirmed your cluster is successfully running CDH 5.



Important:

If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients as well.

Step 7: Set the Sticky Bit

For security reasons Cloudera strongly recommends you set the sticky bit on directories if you have not already done so.

The sticky bit prevents anyone except the superuser, directory owner, or file owner from deleting or moving the files within a directory. (Setting the sticky bit for a file has no effect.) Do this for directories such as `/tmp`. (For instructions on creating `/tmp` and setting its permissions, see [these instructions](#)).

Step 8: Upgrade Components



Note:

- For important information on new and changed components, see the [CDH 5 Release Notes](#). To see whether there is a new version of a particular component in CDH 5, check the [CDH Version and Packaging Information](#).
- Cloudera recommends that you regularly update the software on each system in the cluster (for example, on a RHEL-compatible system, regularly run `yum update`) to ensure that all the dependencies for any given component are up to date. (If you have not been in the habit of doing this, be aware that the command may take a while to run the first time you use it.)

CDH 5 Components

Use the following sections to install or upgrade CDH 5 components:

- [Crunch Installation](#)
- [Flume Installation](#)
- [HBase Installation](#)
- [HCatalog Installation](#)

- [Hive Installation](#)
- [HttpFS Installation](#)
- [Hue Installation](#)
- [Impala Installation](#)
- [KMS Installation and Upgrade](#)
- [Mahout Installation](#)
- [Oozie Installation](#)
- [Pig Installation](#)
- [Search Installation](#)
- [Sentry Installation](#)
- [Snappy Installation](#)
- [Spark Installation](#)
- [Sqoop 1 Installation](#)
- [Sqoop 2 Installation](#)
- [Whirr Installation](#)
- [ZooKeeper Installation](#)

See also the instructions for [installing or updating LZO](#).

Step 9: Apply Configuration File Changes if Necessary



Important: Configuration files

- If you install a newer version of a package that is already on the system, configuration files that you have modified will remain intact.
- If you uninstall a package, the package manager renames any configuration files you have modified from `<file>` to `<file>.rpmsave`. If you then re-install the package (probably to install a new version) the package manager creates a new `<file>` with applicable defaults. You are responsible for applying any changes captured in the original configuration file to the new configuration file. In the case of Ubuntu and Debian upgrades, you will be prompted if you have made changes to a file for which there is a new version. For details, see [Automatic handling of configuration files by dpkg](#).

For example, if you have modified your `zoo.cfg` configuration file (`/etc/zookeeper/zoo.cfg`), the upgrade renames and preserves a copy of your modified `zoo.cfg` as `/etc/zookeeper/zoo.cfg.rpmsave`. If you have not already done so, you should now compare this to the new `/etc/zookeeper/conf/zoo.cfg`, resolve differences, and make any changes that should be carried forward (typically where you have changed property value defaults). Do this for each component you upgrade.

Upgrading from a Release Lower than CDH 5.4.0 to the Latest Release



Important:

Use the instructions on this page to upgrade only from a **CDH 5 release earlier than CDH 5.4.0**.

To upgrade from other releases, see the following topics:

- [Upgrading from CDH 5.4.0 or Higher to the Latest Release](#) on page 186
- [Upgrading from CDH 4 to CDH 5](#) on page 167

Upgrading Unmanaged CDH Using the Command Line

Step 1: Prepare the Cluster for the Upgrade



Important: Before you begin, read the following topics, which contain important upgrade information:

- [Upgrading from an Earlier CDH 5 Release to the Latest Release](#) on page 184
- [Before Upgrading to the Latest Release of CDH](#) on page 185

1. Put the NameNode into safe mode and save the `fsimage`:

a. Put the NameNode (or active NameNode in an HA configuration) into safe mode:

```
$ sudo -u hdfs hdfs dfsadmin -safemode enter
```

b. Run a `saveNamespace` operation:

```
$ sudo -u hdfs hdfs dfsadmin -saveNamespace
```

This results in a new `fsimage` written with no edit log entries.

c.

2. Shut down Hadoop services across your entire cluster by running the following command on every host in your cluster:

```
$ for x in `cd /etc/init.d ; ls hadoop-*` ; do sudo service $x stop ; done
```

3. Check each host to make sure that there are no processes running as the `hdfs`, `yarn`, `mapred` or `httpfs` users from root:

```
# ps -aef | grep java
```

4. Ensure that the NameNode service is not running, and then back up the HDFS metadata on the NameNode machine, as follows.



Note: Cloudera recommends backing up HDFS metadata on a regular basis, as well as before a major upgrade.

a. Find the location of your `dfs.namenode.name.dir`. For example:

```
$ grep -C1 dfs.namenode.name.dir /etc/hadoop/conf/hdfs-site.xml
<property> <name>dfs.namenode.name.dir</name> <value>/mnt/hadoop/hdfs/name</value>
</property>
```

b. Back up the directory. The path inside the `<value>` XML element is the path to your HDFS metadata. If you see a comma-separated list of paths, you do not need to back up all of them; they store the same data. Back up the first directory by using the following commands:

```
$ cd /mnt/hadoop/hdfs/name
# tar -cvf /root/nn_backup_data.tar .
./
./current/
./current/fsimage
./current/fstime
./current/VERSION
./current/edits
./image/
./image/fsimage
```




Important: If you see a file containing the word *lock*, the NameNode is probably still running. Re-run the procedure, beginning at step 1, and make sure that the NameNode service is not running before backing up the HDFS metadata.

Step 2: If Necessary, Download the CDH 5 "1-click" Package on Each of the Hosts in the Cluster

Before you begin: Check whether you have the CDH 5 "1-click" repository installed, and proceed as indicated.

Table 5: Checking for the 1-click Repository

Operating System	Command to Run	Results and Actions
RHEL-compatible	<code>rpm -q cdh5-repository</code>	If the command returns, <code>cdh5-repository-1-0</code> , the 1-click repository is installed. Skip to Step 3 . If the command returns package <code>cdh5-repository</code> is not installed, go to the 1-click instructions for your OS: <ul style="list-style-type: none"> • RHEL • SLES • Ubuntu and Debian
Ubuntu and Debian	<code>dpkg -l grep cdh5-repository</code>	

On RHEL-compatible systems:

1. Download the CDH 5 "1-click Install" package (or RPM).

Click the appropriate RPM and **Save File** to a directory with write access (for example, your home directory).

OS Version	Link to CDH 5 RPM
RHEL/CentOS/Oracle 5	RHEL/CentOS/Oracle 5 link
RHEL/CentOS/Oracle 6	RHEL/CentOS/Oracle 6 link
RHEL/CentOS/Oracle 7	RHEL/CentOS/Oracle 7 link

2. Install the RPM for all RHEL versions:

```
$ sudo yum --nogpgcheck localinstall cloudera-cdh-5-0.x86_64.rpm
```

For instructions on how to add a CDH 5 yum repository or build your own CDH 5 yum repository, see [Installing CDH 5 On Red Hat-compatible systems](#).



Note: Clean repository cache.

Before proceeding, clean cached packages and headers to ensure your system repos are up-to-date:

```
sudo yum clean all
```

On SLES systems:

1. Download the CDH 5 "1-click Install" package.

Upgrading Unmanaged CDH Using the Command Line

Download the [RPM file](#), choose **Save File**, and save it to a directory to which you have write access (for example, your home directory).

2. Install the RPM:

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

3. Update your system package index by running the following:

```
$ sudo zypper refresh
```

```
$ sudo rpm -i cloudera-cdh-5-0.x86_64.rpm
```

For instructions on how to add a repository or build your own repository, see [Installing CDH 5 on SLES Systems](#).



Note: Clean repository cache.

Before proceeding, clean cached packages and headers to ensure your system repos are up-to-date:

```
sudo zypper clean --all
```

On Ubuntu and Debian systems:

1. Download the CDH 5 "1-click Install" package:

OS Version	Package Link
Jessie	Jessie package
Wheezy	Wheezy package
Precise	Precise package
Trusty	Trusty package

2. Install the package by doing one of the following:

- Choose **Open with** in the download window to use the package manager.
- Choose **Save File**, save the package to a directory to which you have write access (for example, your home directory), and install it from the command line. For example:

```
sudo dpkg -i cdh5-repository_1.0_all.deb
```

For instructions on how to add a repository or build your own repository, see the instructions on [installing CDH 5 on Ubuntu and Debian systems](#).



Important: Clean cached packages and headers to ensure that your system repos are up-to-date:

```
sudo apt-get update
```

Step 3: Upgrade the Packages on the Appropriate Hosts

Upgrade [MRv1](#), [YARN](#), or both. Although you can install and configure both MRv1 and YARN, you should not run them both on the same set of hosts at the same time.

If you are using [HA for the NameNode](#), do not install `hadoop-hdfs-secondarynamenode`

Before upgrading MRv1 or YARN: (Optionally) add a repository key on each system in the cluster, if you have not already done so. Add the Cloudera Public GPG Key to your repository by executing one of the following commands:

- **For Red Hat/CentOS/Oracle 5 systems:**

```
$ sudo rpm --import
https://archive.cloudera.com/cdh5/redhat/5/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **For Red Hat/CentOS/Oracle 6 systems:**

```
$ sudo rpm --import
https://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **For all SLES systems:**

```
$ sudo rpm --import
https://archive.cloudera.com/cdh5/sles/11/x86_64/cdh/RPM-GPG-KEY-cloudera
```

- **For Ubuntu Precise systems:**

```
$ curl -s
https://archive.cloudera.com/cdh5/ubuntu/precise/amd64/cdh/archive.key
| sudo apt-key add -
```

- **For Debian Wheezy systems:**

```
$ curl -s
https://archive.cloudera.com/cdh5/debian/wheezy/amd64/cdh/archive.key
| sudo apt-key add -
```

Step 3a: If you are using MRv1, upgrade the MRv1 packages on the appropriate hosts.

Skip this step if you are using [YARN](#) exclusively.

1. Install and deploy ZooKeeper as described in [ZooKeeper Installation](#). Cloudera recommends that you install (or update) and start a ZooKeeper cluster, and ZooKeeper is required if you are deploying high availability (HA) for the NameNode or JobTracker.
2. Install each daemon package on the appropriate systems, as follows.

Where to install	Install commands
JobTracker host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-jobtracker</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-jobtracker</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-jobtracker</code>
NameNode host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>

Where to install	Install commands
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>
All cluster hosts except the JobTracker, NameNode, and Secondary (or Standby) NameNode hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-datanode</code>
All client hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-client</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-client</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-client</code>

Step 3b: If you are using YARN, upgrade the YARN packages on the appropriate hosts.

Skip this step if you are using [MRv1](#) exclusively.

1. Install and deploy ZooKeeper as described in [ZooKeeper Installation](#). Cloudera recommends that you install (or update) and start a ZooKeeper cluster, and ZooKeeper is required if you are deploying high availability (HA) for the NameNode or JobTracker.
2. Install each type of daemon package on the appropriate systems(s), as follows.

Where to install	Install commands
Resource Manager host (analogous to MRv1 JobTracker) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-yarn-resourcemanager</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-yarn-resourcemanager</code>

Where to install	Install commands
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-yarn-resourcemanager</code>
NameNode host running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-namenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-namenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-namenode</code>
Secondary NameNode host (if used) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-hdfs-secondarynamenode</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-hdfs-secondarynamenode</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-hdfs-secondarynamenode</code>
All cluster hosts except the Resource Manager (analogous to MRv1 TaskTrackers) running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-yarn-nodemanager hadoop-hdfs-datanode hadoop-mapreduce</code>
One host in the cluster running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
<i>Ubuntu or Debian</i>	<code>\$ sudo apt-get update; sudo apt-get install hadoop-mapreduce-historyserver hadoop-yarn-proxyserver</code>
All client hosts, running:	
<i>Red Hat/CentOS compatible</i>	<code>\$ sudo yum clean all; sudo yum install hadoop-client</code>

Where to install	Install commands
<i>SLES</i>	<code>\$ sudo zypper clean --all; sudo zypper install hadoop-client</code>
<i>Ubuntu or Debian</i>	<code>sudo apt-get update; sudo apt-get install hadoop-client</code>

**Note:**

The `hadoop-yarn` and `hadoop-hdfs` packages are installed on each system automatically as dependencies of the other packages.

Step 4: In an HA Deployment, Upgrade and Start the JournalNodes

1. Install the JournalNode daemons on each of the machines where they will run.

To install JournalNode on RHEL-compatible systems:

```
$ sudo yum install hadoop-hdfs-journalnode
```

To install JournalNode on Ubuntu and Debian systems:

```
$ sudo apt-get install hadoop-hdfs-journalnode
```

To install JournalNode on SLES systems:

```
$ sudo zypper install hadoop-hdfs-journalnode
```

2. Start the JournalNode daemons on each of the machines where they will run:

```
sudo service hadoop-hdfs-journalnode start
```

Wait for the daemons to start before proceeding to the next step.

**Important:**

In an HA deployment, the JournalNodes must be up and running CDH 5 before you proceed.

Step 5: Upgrade the HDFS Metadata

The steps for upgrading HDFS metadata differ for HA and non-HA deployments.

Section 5a: Upgrade the HDFS Metadata for HA Deployments

1. Make sure that the JournalNodes have been upgraded to CDH 5 and are up and running.
2. Run the following command on the *active NameNode only*:

```
$ sudo service hadoop-hdfs-namenode upgrade
```

**Warning:**

In an HDFS HA deployment, it is critically important that you do this on only one NameNode.

3. Monitor the progress of the metadata upgrade by running the following:

```
$ sudo tail -f /var/log/hadoop-hdfs/hadoop-hdfs-namenode-<hostname>.log
```

Look for a line that confirms the upgrade is complete, such as:

```
/var/lib/hadoop-hdfs/cache/hadoop/dfs/<name> is complete.
```

The NameNode upgrade process can take a while, depending on the number of files.

4. Wait for NameNode to exit safe mode, and then restart the standby NameNode.

- If Kerberos is enabled:

```
$ kinit -kt /path/to/hdfs.keytab hdfs/<fully.qualified.domain.name@YOUR-REALM.COM> &&
hdfs namenode -bootstrapStandby
```

```
$ sudo service hadoop-hdfs-namenode start
```

- If Kerberos is not enabled:

```
$ sudo -u hdfs hdfs namenode -bootstrapStandby
$ sudo service hadoop-hdfs-namenode start
```

5. Start the DataNodes by running the following command on each DataNode:

```
$ sudo service hadoop-hdfs-datanode start
```

Section 5b: Upgrade the HDFS Metadata for Non-HA Deployments

1. Run the following command on the NameNode:

```
$ sudo service hadoop-hdfs-namenode upgrade
```

2. Monitor the progress of the metadata upgrade by running the following:

```
$ sudo tail -f /var/log/hadoop-hdfs/hadoop-hdfs-namenode-<hostname>.log
```

Look for a line that confirms the upgrade is complete, such as:

```
/var/lib/hadoop-hdfs/cache/hadoop/dfs/<name> is complete.
```

The NameNode upgrade process can take a while, depending on the number of files.

3. Start the DataNodes by running the following command on each DataNode:

```
$ sudo service hadoop-hdfs-datanode start
```

4. Wait for NameNode to exit safe mode, and then start the secondary NameNode:

1. To check that the NameNode has exited safe mode, look for messages in the log file, or the NameNode's web interface, that say "...no longer in safe mode."

2. To start the secondary NameNode, enter the following command on the secondary NameNode host:

```
$ sudo service hadoop-hdfs-secondarynamenode start
```

Step 6: Start MapReduce (MRv1) or YARN

You are now ready to start and test [MRv1](#) or [YARN and the MapReduce JobHistory Server](#).

**Important:**

Do not run MRv1 and YARN on the same set of hosts at the same time. This degrades performance and can result in an unstable cluster deployment. Steps 6a and 6b are mutually exclusive.

Step 6a: Start MRv1

1. Start each TaskTracker:

```
$ sudo service hadoop-0.20-mapreduce-tasktracker start
```

2. Start each JobTracker:

```
$ sudo service hadoop-0.20-mapreduce-jobtracker start
```

3. Verify that the JobTracker and TaskTracker started properly:

```
$ sudo jps | grep Tracker
```

If the permissions of directories are not configured correctly, the JobTracker and TaskTracker processes start and immediately fail. If this happens, check the JobTracker and TaskTracker logs and set the permissions correctly.

4. Verify basic cluster operation for MRv1.

Before running production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.

**Important:**

For important cluster configuration information, see [Deploying MapReduce v1 \(MRv1\) on a Cluster](#).

a. Create a home directory on HDFS for user joe:

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/joe
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Perform steps a through f as user joe.

b. Make a directory in HDFS called input and copy some XML files into it by running the following commands:

```
$ hadoop fs -mkdir input
$ hadoop fs -put /etc/hadoop/conf/*.xml input
$ hadoop fs -ls input
Found 3 items:
-rw-r--r-- 1 joe supergroup 1348 2012-02-13 12:21 input/core-site.xml
-rw-r--r-- 1 joe supergroup 1913 2012-02-13 12:21 input/hdfs-site.xml
-rw-r--r-- 1 joe supergroup 1001 2012-02-13 12:21 input/mapred-site.xml
```

c. Run an example Hadoop job to grep with a regular expression in your input data:

```
$ /usr/bin/hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples.jar grep input
output 'dfs[a-z.]+'
```

d. After the job completes, find the output in the HDFS directory named output which you specified to Hadoop:

```
$ hadoop fs -ls
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x - joe supergroup 0 2009-08-18 18:38 /user/joe/output
```


e. List the output files:

```
$ hadoop fs -ls output
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-02-25 10:33 /user/joe/output/_logs
-rw-r--r-- 1 joe supergroup 1068 2009-02-25 10:33 /user/joe/output/part-00000
-rw-r--r- 1 joe supergroup 0 2009-02-25 10:33 /user/joe/output/_SUCCESS
```

f. Read the results in the output file; for example:

```
$ hadoop fs -cat output/part-00000 | head
1 dfs.datanode.data.dir
1 dfs.namenode.checkpoint.dir
1 dfs.namenode.name.dir
1 dfs.replication
1 dfs.safemode.extension
1 dfs.safemode.min.datanodes
```

This confirms your cluster is successfully running CDH 5.



Important:

If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients.

Step 6b: Start MapReduce with YARN

1. If you have not already done so, create directories and set the correct permissions.



Note: For more information about YARN configuration and permissions, see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

a. Create a history directory and set permissions; for example:

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/history
$ sudo -u hdfs hadoop fs -chmod -R 1777 /user/history
$ sudo -u hdfs hadoop fs -chown yarn /user/history
```

b. Create the `/var/log/hadoop-yarn` directory and set ownership:

```
$ sudo -u hdfs hadoop fs -mkdir -p /var/log/hadoop-yarn
$ sudo -u hdfs hadoop fs -chown yarn:mapred /var/log/hadoop-yarn
```

You create this directory because it is the parent of `/var/log/hadoop-yarn/apps`, which is explicitly configured in the `yarn-site.xml`.

c. Verify the directory structure, ownership, and permissions:

```
$ sudo -u hdfs hadoop fs -ls -R /
```

You should see:

```
drwxrwxrwt - hdfs supergroup 0 2012-04-19 14:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2012-05-31 10:26 /user
drwxrwxrwt - yarn supergroup 0 2012-04-19 14:31 /user/history
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var
drwxr-xr-x - hdfs supergroup 0 2012-05-31 15:31 /var/log
drwxr-xr-x - yarn mapred 0 2012-05-31 15:31 /var/log/hadoop-yarn
```

2. Start YARN, and start the ResourceManager and NodeManager services:



Important:

Always start ResourceManager before starting NodeManager services.

- a. On the ResourceManager system, run the following command:

```
$ sudo service hadoop-yarn-resourcemanager start
```

- b. On each NodeManager system (typically the same ones where DataNode service runs):

```
$ sudo service hadoop-yarn-nodemanager start
```

3. Start the MapReduce JobHistory Server:

- a. On the MapReduce JobHistory Server system, run the following command:

```
$ sudo service hadoop-mapreduce-historyserver start
```

- b. For each user who will be submitting MapReduce jobs using YARN, or running Pig, Hive, or Sqoop 1 in a YARN installation, set the HADOOP_MAPRED_HOME environment variable as follows:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

4. Before running production jobs, verify basic cluster operation by running an example from the Apache Hadoop web site.



Note:

For important configuration information, see [Deploying MapReduce v2 \(YARN\) on a Cluster](#).

- a. Create a home directory for user joe:

```
$ sudo -u hdfs hadoop fs -mkdir -p /user/joe  
$ sudo -u hdfs hadoop fs -chown joe /user/joe
```

Perform the remaining steps as the user joe.

- b. Make a directory in HDFS called input and copy XML files to it by running the following commands in pseudo-distributed mode:

```
$ hadoop fs -mkdir input  
$ hadoop fs -put /etc/hadoop/conf/*.xml input  
$ hadoop fs -ls input  
Found 3 items:  
-rw-r--r-- 1 joe supergroup 1348 2012-02-13 12:21 input/core-site.xml  
-rw-r--r-- 1 joe supergroup 1913 2012-02-13 12:21 input/hdfs-site.xml  
-rw-r--r-- 1 joe supergroup 1001 2012-02-13 12:21 input/mapred-site.xml
```

- c. Set HADOOP_MAPRED_HOME for user joe:

```
$ export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

- d. Run an example Hadoop job to grep with a regular expression in your input data:

```
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar grep input output23  
'dfs[a-z.]+'
```

- e. After the job completes, find the output in the HDFS directory named `output23`, which you specified to Hadoop:

```
$ hadoop fs -ls
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-08-18 18:36 /user/joe/input
drwxr-xr-x - joe supergroup 0 2009-08-18 18:38 /user/joe/output23
```

- f. List the output files:

```
$ hadoop fs -ls output23
Found 2 items
drwxr-xr-x - joe supergroup 0 2009-02-25 10:33 /user/joe/output23/_SUCCESS
-rw-r--r-- 1 joe supergroup 1068 2009-02-25 10:33 /user/joe/output23/part-r-00000
```

- g. Read the results in the output file:

```
$ hadoop fs -cat output23/part-r-00000 | head
1 dfs.safemode.min.datanodes
1 dfs.safemode.extension
1 dfs.replication
1 dfs.permissions.enabled
1 dfs.namenode.name.dir
1 dfs.namenode.checkpoint.dir
1 dfs.datanode.data.dir
```

This confirms that your cluster is successfully running CDH 5.



Important:

If you have client hosts, make sure you also update them to CDH 5, and upgrade the [components](#) running on those clients as well.

Step 7: Set the Sticky Bit

For security reasons, Cloudera strongly recommends you set the sticky bit on directories if you have not already done so.

The sticky bit prevents anyone except the superuser, directory owner, or file owner from deleting or moving the files within a directory. (Setting the sticky bit for a file has no effect.) Do this for directories such as `/tmp`. (For instructions on creating `/tmp` and setting its permissions, see [Create the /tmp Directory](#)).

Step 8: Upgrade Components

Cloudera recommends that you regularly update the software on each system in the cluster (for example, on a RHEL-compatible system, regularly run `yum update`) to ensure that all the dependencies for any given component are up to date. If you have not been doing this, the command may take a while to run the first time you use it.



Note:

- For important information on new and changed components, see the [CDH 5 Release Notes](#). To see whether there is a new version of a particular component in CDH 5, check the [CDH Version and Packaging Information](#).

CDH 5 Components

Use the following sections to install or upgrade CDH 5 components:

- [Crunch Installation](#)
- [Flume Installation](#)

Upgrading Unmanaged CDH Using the Command Line

- [HBase Installation](#)
- [HCatalog Installation](#)
- [Hive Installation](#)
- [HttpFS Installation](#)
- [Hue Installation](#)
- [Impala Installation](#)
- [KMS Installation and Upgrade](#)
- [Mahout Installation](#)
- [Oozie Installation](#)
- [Pig Installation](#)
- [Search Installation](#)
- [Sentry Installation](#)
- [Snappy Installation](#)
- [Spark Installation](#)
- [Sqoop 1 Installation](#)
- [Sqoop 2 Installation](#)
- [Whirr Installation](#)
- [ZooKeeper Installation](#)

See also the instructions for [installing or updating LZO](#).

Step 9: Apply Configuration File Changes if Required



Important: Configuration files

- If you install a newer version of a package that is already on the system, configuration files that you have modified will remain intact.
- If you uninstall a package, the package manager renames any configuration files you have modified from `<file>` to `<file>.rpmsave`. If you then re-install the package (probably to install a new version) the package manager creates a new `<file>` with applicable defaults. You are responsible for applying any changes captured in the original configuration file to the new configuration file. In the case of Ubuntu and Debian upgrades, you will be prompted if you have made changes to a file for which there is a new version. For details, see [Automatic handling of configuration files by dpkg](#).

For example, if you have modified your `zoo.cfg` configuration file (`/etc/zookeeper/zoo.cfg`), the upgrade renames and preserves a copy of your modified `zoo.cfg` as `/etc/zookeeper/zoo.cfg.rpmsave`. If you have not already done so, you should now compare this to the new `/etc/zookeeper/conf/zoo.cfg`, resolve differences, and make any changes that should be carried forward (typically where you have changed property value defaults). Do this for each component you upgrade.

Step 10: Finalize the HDFS Metadata Upgrade



Important: Once you have finalized the upgrade, you cannot roll back to a previous version of HDFS. See [Rolling Back a CDH 4-to-CDH 5 Upgrade](#) on page 117.

To finalize the HDFS metadata upgrade, do the following:

1. Make sure that the CDH 5 upgrade has succeeded and everything is running as expected. You can wait days or even weeks to verify a successful upgrade before finalizing it.

Before finalizing, run important workloads and ensure that they are successful. Once you have finalized the upgrade, you cannot roll back to a previous version of HDFS without using backups.

**Note:**

- If you need to restart the NameNode during this period (after having begun the upgrade process, but before you have run `finalizeUpgrade`), restart your NameNode without the `-upgrade` option.
- Verifying that you are ready to finalize the upgrade can take a long time. Make sure you have enough free disk space, keeping in mind that the following behavior continues until the upgrade is finalized:
 - Deleting files does not free up disk space.
 - Using the balancer causes all moved replicas to be duplicated.
 - All on-disk data representing the NameNodes metadata is retained, which could more than double the amount of space required on the NameNode and JournalNode disks.

2. Finalize the HDFS metadata upgrade by using one of the following commands, depending on whether Kerberos is enabled (see [Enabling Kerberos Authentication for Hadoop Using the Command Line](#)).



Important: In an HDFS HA deployment, make sure that both the NameNodes and all of the JournalNodes are up and functioning normally before you proceed.

- If Kerberos is enabled:

```
$ kinit -kt /path/to/hdfs.keytab hdfs/<fully.qualified.domain.name@YOUR-REALM.COM> &&
hdfs dfsadmin -finalizeUpgrade
```

- If Kerberos is not enabled:

```
$ sudo -u hdfs hdfs dfsadmin -finalizeUpgrade
```

After the metadata upgrade completes, the `previous/` and `blocksBeingWritten/` directories in the DataNode data directories are not cleared until the DataNodes are restarted.

Troubleshooting: If You Missed the HDFS Metadata Upgrade Steps

If you skipped [Step 5: Upgrade the HDFS Metadata](#) on page 206, HDFS will not start; the metadata upgrade is required for all upgrades to CDH 5.4.0 and higher from any earlier release. You will see errors such as the following:

```
2014-10-16 18:36:29,112 WARN org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
Encountered exception loading fsimage
    java.io.IOException:File system image contains an old layout version -55.An
upgrade to version -59 is required.
    Please restart NameNode with the "-rollingUpgrade started" option if a rolling
upgrade is already started; or restart NameNode with the "-upgrade"
option to start a new upgrade.
        at
org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:231)
        at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:994)
        at
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:726)
        at
org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:529)
        at
```

```

    org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:585)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:751)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:735)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1410)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1476)
2014-10-16 18:36:29,126 INFO org.mortbay.log: Stopped
HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:50070
2014-10-16 18:36:29,127 WARN org.apache.hadoop.http.HttpServer2: HttpServer
Acceptor: isRunning is false. Rechecking.
2014-10-16 18:36:29,127 WARN org.apache.hadoop.http.HttpServer2: HttpServer
Acceptor: isRunning is false
2014-10-16 18:36:29,127 INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl:
Stopping NameNode metrics system...
2014-10-16 18:36:29,128 INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl:
NameNode metrics system stopped.
2014-10-16 18:36:29,128 INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl:
NameNode metrics system shutdown complete.
2014-10-16 18:36:29,128 FATAL org.apache.hadoop.hdfs.server.namenode.NameNode:
Exception in namenode join
java.io.IOException: File system image contains an old layout version -55.An
upgrade to version -59 is required.
Please restart NameNode with the "-rollingUpgrade started" option if a rolling
upgrade is already
started; or restart NameNode with the "-upgrade" option to start a new upgrade.
        at
    org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:231)
        at
    org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:994)
        at
    org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:726)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:529)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:585)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:751)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:735)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1410)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1476)
2014-10-16 18:36:29,130 INFO org.apache.hadoop.util.ExitUtil: Exiting with status
1
2014-10-16 18:36:29,132 INFO org.apache.hadoop.hdfs.server.namenode.NameNode:
SHUTDOWN_MSG:

```

To recover, proceed as follows:

1. Make sure you have completed all the necessary preceding steps ([Step 1: Prepare the Cluster for the Upgrade](#) on page 200 through [Step 4: In an HA Deployment, Upgrade and Start the JournalNodes](#) on page 206; or [Step 1: Prepare the Cluster for the Upgrade](#) on page 200 through [Step 3: Upgrade the Packages on the Appropriate Hosts](#) on page 202 if this is not an HA deployment).
2. Starting with [Step 5: Upgrade the HDFS Metadata](#) on page 206, complete all the remaining steps through [Step 10: Finalize the HDFS Metadata Upgrade](#) on page 212.

Upgrading Host Operating Systems in a CDH Cluster

This topic describes the steps to upgrade the operating system (OS) on hosts in an active CDH cluster.

Prerequisites

1. If you are upgrading CDH or Cloudera Manager as well as the OS, upgrade the OS first.
2. Read the release notes of the specific CDH version to ensure that the new OS version is supported. In most cases, upgrading to a minor version of the OS is supported. Read the release note of the new OS release to check for new default settings or changes in behavior. (For example, Transparent Huge Pages is on by default in some versions of RHEL 6).
3. If any files have a replication factor lower than the global default, verify that bringing down any node does not make the file unavailable.

Upgrading Hosts

For each host in the cluster, do the following:

1. If the host runs a DataNode service, determine whether it needs to be decommissioned. See [Deciding Whether to Decommission DataNodes](#) on page 215.
2. Stop all Hadoop-related services on the host.
3. Take the host offline (for example, switch to single-user mode or restart the host to boot off the network).
4. Upgrade the OS partition, leaving the data partitions (for example, `dfs.data.dir`) unchanged.
5. Bring the host back online.
6. If the host is decommissioned, recommission it.
7. Verify in the NameNode UI (or Cloudera Manager) that the host is healthy and all services are running.

Upgrading Hosts With High Availability Enabled

If you have enabled high availability for the NameNode or JobTracker, follow this procedure:

1. Stop the backup NameNode or JobTracker.
2. Upgrade the OS partition.
3. Start the services and ensure that they are running properly.
4. Fail over to the backup NameNode or JobTracker.
5. Upgrade the primary NameNode or JobTracker.
6. Bring the primary NameNode or JobTracker back online.
7. Reverse the failover.

Upgrading Hosts Without High Availability Enabled

If you have not enabled high availability, upgrading a primary host causes an outage for that service. The procedure to upgrade is the same as upgrading a secondary host, except that you must decommission and recommission the host. When upgrading hosts that are part of a ZooKeeper quorum, ensure that the majority of the quorum is available. Cloudera recommends that you upgrade only one host at a time.

Deciding Whether to Decommission DataNodes

When a DataNode is decommissioned, the NameNode ensures that every block from the DataNode is still available across the cluster as specified by the replication factor. This procedure involves copying blocks off the DataNode in small batches. In cases where a DataNode has several thousand blocks, decommissioning takes several hours.

When a DataNode is turned off without being decommissioned:

- The NameNode marks the DataNode as dead after a default of 10m 30s (controlled by `dfs.heartbeat.interval` and `dfs.heartbeat.recheck.interval`).

Upgrading Host Operating Systems in a CDH Cluster

- The NameNode schedules the missing replicas to be placed on other DataNodes.
- When the DataNode comes back online and reports to the NameNode, the NameNode schedules blocks to be copied to it while other nodes are decommissioned or when new files are written to HDFS.

If the OS upgrade procedure is quick (for example, under 30 mins per node), do not decommission the DataNode.

Speed up the decommissioning of a DataNode by increasing values for these properties:

- `dfs.max-repl-streams`: The number of simultaneous streams to copy data.
- `dfs.balance.bandwidthPerSec`: The maximum amount of bandwidth that each DataNode can utilize for balancing, in bytes per second.
- `dfs.namenode.replication.work.multiplier.per.iteration`: NameNode configuration requiring a restart, defaults to 2 but can be raised to 10 or higher.

This determines the total amount of block transfers to begin in parallel at a DataNode for replication, when such a command list is being sent over a DataNode heartbeat by the NameNode. The actual number is obtained by multiplying this value by the total number of live nodes in the cluster. The result number is the number of blocks to transfer immediately, per DataNode heartbeat.

For more information, see [Decommissioning and Recommissioning Hosts](#).