cloudera®

Cloudera Installation

Important Notice
© 2010-2021 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. If this documentation includes code, including but not limited to, code examples, Cloudera makes this available to you under the terms of the Apache License, Version 2.0, including any required notices. A copy of the Apache License Version 2.0, including any notices, is included herein. A copy of the Apache License Version 2.0 can also be found here: https://opensource.org/licenses/Apache-2.0

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.
395 Page Mill Road
Palo Alto, CA 94306
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-362-0488
www.cloudera.com

Release Information

Version: Cloudera Enterprise 5.16.x
Date: August 3, 2021

Table of Contents

Cloudera Installation Guide....8

Before You Install....9

Storage Space Planning for Cloudera Manager....9

Cloudera Manager Server....9

Cloudera Management Service....10

Cloudera Navigator....15

Cluster Lifecycle Management with Cloudera Manager....17

Configure Network Names....20

Required Privileges for Package-based Installations of CDH....21

Required Tomcat Directories....22

Ports....22

Ports Used by Cloudera Manager and Cloudera Navigator....23

Ports Used by Cloudera Navigator Encryption....28

Ports Used by Components of CDH....28

Ports Used by Impala....35

Ports Used by Cloudera Search....37

Ports Used by DistCp....37

Ports Used by Third-Party Components....38

Ports Used by Apache Flume and Apache Solr....39

Recommended Cluster Hosts and Role Distribution....39

CDH Cluster Hosts and Role Assignments....39

Allocating Hosts for Key Trustee Server and Key Trustee KMS....43

Custom Installation Solutions....43

Introduction to Parcels....44

Understanding Package Management....44

Using an Internal Parcel Repository....45

Using an Internal Package Repository....48

Manually Install Cloudera Software Packages....50

Installing Cloudera Software Using Tarballs....52

Creating Virtual Images of Cluster Hosts....63

Configuring a Custom Java Home Location....66

Creating a CDH Cluster Using a Cloudera Manager Template....66

Configuring Single User Mode....72

Installing Cloudera Manager, CDH, and Managed Services....82

Step 1: Configure a Repository for Cloudera Manager....82
Step 2: Install Java Development Kit....84
Requirements....84
Installing Oracle JDK Using Cloudera Manager....85
Manually Installing Oracle JDK....85
Manually Installing OpenJDK....85
Step 3: Install Cloudera Manager Server....86
Install the Cloudera Manager Server Packages....86
Install and Configure Databases....87
Step 4: Install and Configure Databases....87
Required Databases....88
Installing and Configuring Databases....88
Install and Configure MariaDB for Cloudera Software....88
Install and Configure MySQL for Cloudera Software....94
Install and Configure PostgreSQL for Cloudera Software....100
Install and Configure Oracle Database for Cloudera Software....104
Configuring an External Database for Sqoop 2....114
Step 5: Set up the Cloudera Manager Database....115
Syntax for `scm_prepare_database.sh`....115
Preparing the Cloudera Manager Server Database....117
Installing CDH....118
Step 6: Install CDH and Other Software....118
Welcome to Cloudera Manager....119
Thank you for choosing Cloudera Manager and CDH....119
Specify hosts for your CDH cluster installation....119
Select Repository....120
Accept JDK License....121
Single User Mode....121
Enter Login Credentials....121
Install Agents....122
Install Parcels....122
Inspect hosts for correctness....122
Step 7: Set Up a Cluster Using the Wizard....122
Select Services....122
Assign Roles....123
Setup Database....123
Review Changes....123
First Run Command....123
Congratulations!....123

Installing the Cloudera Navigator Data Management Component....124

Installing Cloudera Navigator Encryption Components....127

Installing Cloudera Navigator Key Trustee Server....127
Prerequisites....127
Setting Up an Internal Repository....127
Installing Key Trustee Server....127
Securing Key Trustee Server Host....130
Leveraging Native Processor Instruction Sets....131
Initializing Key Trustee Server....131
Installing Cloudera Navigator Key HSM....132
Prerequisites....132
Setting Up an Internal Repository....132
Installing Navigator Key HSM....132
Installing Key Trustee KMS....133
Setting Up an Internal Repository....133
Installing Key Trustee KMS Using Parcels....133
Installing Key Trustee KMS Using Packages....133
Post-Installation Configuration....134
Installing Navigator HSM KMS Backed by Thales HSM 134
Client Prerequisites....134
Setting Up an Internal Repository....135
Installing Navigator HSM KMS Backed by Thales HSM Using Parcels....135
Installing Navigator HSM KMS Backed by Thales HSM Using Packages....135
Installing Navigator HSM KMS Backed by Luna HSM 135
Client Prerequisites....136
Setting Up an Internal Repository....136
Installing Navigator HSM KMS Backed by Luna HSM Using Parcels....136
Installing Navigator HSM KMS Backed by Luna HSM Using Packages....136
Post-Installation Configuration....136
Installing Cloudera Navigator Encrypt....137
Prerequisites....137
Setting Up an Internal Repository....137
Installing Navigator Encrypt (RHEL-Compatible)....137
Installing Navigator Encrypt (SLES)....139
Installing Navigator Encrypt (Debian or Ubuntu)....140
Post Installation....140
Setting Up TLS for Navigator Encrypt Clients....140
Entropy Requirements....141
Uninstalling and Reinstalling Navigator Encrypt....142

After Installation....144

Deploying Clients....144
Testing the Installation....144
Checking Host Heartbeats....144
Running a MapReduce Job....145

Testing with Hue *145*
Installing the GPL Extras Parcel 145
Migrating from Packages to Parcels 146
Migrating from Parcels to Packages 148
Install CDH and Managed Service Packages *148*
Deactivate Parcels *151*
Restart the Cluster *151*
Remove and Delete Parcels *151*

Troubleshooting Installation Problems 152

Navigator HSM KMS Backed by Thales HSM installation fails 152
Possible Reasons *152*
Possible Solutions *152*
Failed to start server reported by cloudera-manager-installer.bin 152
Possible Reasons *152*
Possible Solutions *152*
Installation interrupted and installer does not restart 152
Possible Reasons *152*
Possible Solutions *152*
Cloudera Manager Server fails to start with MySQL 153
Possible Reasons *153*
Possible Solutions *153*
Agents fail to connect to Server 153
Possible Reasons *153*
Possible Solutions *153*
Cluster hosts do not appear 153
Possible Reasons *153*
Possible Solutions *153*
"Access denied" in install or update wizard 153
Possible Reasons *153*
Possible Solutions *154*
Databases fail to start 154
Possible Reasons *154*
Possible Solutions *154*
Cloudera services fail to start 154
Possible Reasons *154*
Possible Solutions *154*
Activity Monitor displays a status of **BAD** 154
Possible Reasons *155*
Possible Solutions *155*
Activity Monitor fails to start 155
Possible Reasons *155*
Possible Solutions *155*

Create Hive Metastore Database Tables command fails....155
Possible Reasons....155
Possible Solutions....155
Oracle invalid identifier....155
Possible Reasons....155
Possible Solutions....155

Uninstalling Cloudera Manager and Managed Software....156
Uninstalling Cloudera Manager and Managed Software....156
Record User Data Paths....156
Stop all Services....156
Deactivate and Remove Parcels....156
Delete the Cluster....157
Uninstall the Cloudera Manager Server....157
Uninstall Cloudera Manager Agent and Managed Software....157
Remove Cloudera Manager and User Data....159
Uninstalling a CDH Component From a Single Host....160

Appendix: Apache License, Version 2.0....161

Cloudera Installation Guide

This guide provides instructions for installing Cloudera software, including Cloudera Manager, CDH, and other managed services, in a production environment.

For non-production environments (such as testing and proof-of- concept use cases), see Proof-of-Concept Installation Guide for a simplified (but limited) installation procedure.

To install an unmanaged CDH cluster without Cloudera Manager, see Cloudera Command Line Installation Guide.

Important: As of February 1, 2021, all downloads of CDH and Cloudera Manager require a username and password and use a modified URL. You must use the modified URL, including the username and password when downloading the repository contents described below. You may need to upgrade Cloudera Manager to a newer version that uses the modified URLs.

This can affect new installations, upgrades, adding new hosts to a cluster, downloading a new parcel, and adding a new cluster.

For more information, see Updating an existing CDH/Cloudera Manager deployment to access downloads with authentication.

This guide includes the following sections:

Before You Install

Before you install Cloudera Manager, CDH, and other managed services, review CDH 5 and Cloudera Manager 5 Requirements and Supported Versions.

For planning, best practices, and recommendations, review the reference architecture for your environment. For example, for on-premises deployments, review the Cloudera Enterprise Reference Architecture for Bare Metal Deployments (PDF).

The following topics describe additional considerations you should be aware of before beginning an installation:

Storage Space Planning for Cloudera Manager

Minimum Required Role: Full Administrator

Cloudera Manager tracks metrics of services, jobs, and applications in many background processes. All of these metrics require storage. Depending on the size of your organization, this storage can be local or remote, disk-based or in a database, managed by you or by another team in another location.

Most system administrators are aware of common locations like `/var/log/` and the need for these locations to have adequate space. This topic helps you plan for the storage needs and data storage locations used by the Cloudera Manager Server and the Cloudera Management Service to store metrics and data.

Failing to plan for the storage needs of all components of the Cloudera Manager Server and the Cloudera Management Service can negatively impact your cluster in the following ways:

- The cluster might not be able to retain historical operational data to meet internal requirements.
- The cluster might miss critical audit information that was not gathered or retained for the required length of time.
- Administrators might be unable to research past events or health status.
- Administrators might not have historical MR1, YARN, or Impala usage data when they need to reference or report on them later.
- There might be gaps in metrics collection and charts.
- The cluster might experience data loss due to filling storage locations to 100% of capacity. The effects of such an event can impact many other components.

The main theme here is that you must architect your data storage needs well in advance. You must inform your operations staff about your critical data storage locations for each host so that they can provision your infrastructure adequately and back it up appropriately. Make sure to document the discovered requirements in your internal build documentation and run books.

This topic describes both local disk storage and RDBMS storage. This distinction is made both for storage planning and also to inform migration of roles from one host to another, preparing backups, and other lifecycle management events.

The following tables provide details about each individual Cloudera Management service to enable Cloudera Manager administrators to make appropriate storage and lifecycle planning decisions.

Cloudera Manager Server

Table 1: Cloudera Manager Server

Configuration Topic	Cloudera Manager Server Configuration
Default Storage Location	**RDBMS:** Any Supported RDBMS. For more information, see CDH and Cloudera Manager Supported Databases. **Disk:**

Configuration Topic	Cloudera Manager Server Configuration
	Cloudera Manager Server Local Data Storage Directory (`command_storage_path`) on the host where the Cloudera Manager Server is configured to run. This local path is used by Cloudera Manager for storing data, including command result files. Critical configurations are not stored in this location. Default setting: `/var/lib/cloudera-scm-server/`
Storage Configuration Defaults, Minimum, or Maximum	There are no direct storage defaults relevant to this entity.
Where to Control Data Retention or Size	The size of the Cloudera Manager Server database varies depending on the number of managed hosts and the number of discrete commands that have been run in the cluster. To configure the size of the retained command results in the Cloudera Manager Administration Console, select **Administration** > **Settings** and edit the following property: **Command Eviction Age** Length of time after which inactive commands are evicted from the database. Default is two years.
Sizing, Planning & Best Practices	The Cloudera Manager Server database is the most vital configuration store in a Cloudera Manager deployment. This database holds the configuration for clusters, services, roles, and other necessary information that defines a deployment of Cloudera Manager and its managed hosts. Make sure that you perform regular, verified, remotely-stored backups of the Cloudera Manager Server database.

Cloudera Management Service

Table 2: Cloudera Management Service - Activity Monitor Configuration

Configuration Topic	Activity Monitor
Default Storage Location	Any Supported RDBMS. For more information, see CDH and Cloudera Manager Supported Databases.
Storage Configuration Defaults / Minimum / Maximum	Default: 14 Days worth of MapReduce (MRv1) jobs/tasks
Where to Control Data Retention or Size	You control Activity Monitor storage usage by configuring the number of days or hours of data to retain. Older data is purged. To configure data retention in the Cloudera Manager Administration Console: 1. Go the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Activity Monitor or Cloudera Management Service (Service-Wide)**. 4. Select **Category** > **Main**. 5. Locate the following properties or search for them by typing the property name in the **Search** box:

Configuration Topic	Activity Monitor
	Purge Activities Data at This Age In Activity Monitor, purge data about MapReduce jobs and aggregate activities when the data reaches this age in hours. By default, Activity Monitor keeps data about activities for 336 hours (14 days). **Purge Attempts Data at This Age** In the Activity Monitor, purge data about MapReduce attempts when the data reaches this age in hours. Because attempt data can consume large amounts of database space, you might want to purge it more frequently than activity data. By default, Activity Monitor keeps data about attempts for 336 hours (14 days). **Purge MapReduce Service Data at This Age** The number of hours of past service-level data to keep in the Activity Monitor database, such as total slots running. The default is to keep data for 336 hours (14 days). 6. Click **Save Changes** to commit the changes.
Sizing, Planning, and Best Practices	The Activity Monitor only monitors MapReduce jobs, and does not monitor YARN applications. If you no longer use MapReduce (MRv1) in your cluster, the Activity Monitor is not required for Cloudera Manager 5 (or higher) or CDH 5 (or higher). The amount of storage space needed for 14 days worth of MapReduce activities can vary greatly and directly depends on the size of your cluster and the level of activity that uses MapReduce. It might be necessary to adjust and readjust the amount of storage as you determine the "stable state" and "burst state" of the MapReduce activity in your cluster. For example, consider the following test cluster and usage: • A simulated 1000-host cluster, each host with 32 slots • MapReduce jobs with 200 attempts (tasks) per activity (job) Sizing observations for this cluster: • Each attempt takes 10 minutes to complete. • This usage results in roughly 20 thousand jobs a day with approximately 5 million total attempts. • For a retention period of 7 days, this Activity Monitor database required 200 GB.

Table 3: Cloudera Management Service - Service Monitor Configuration

Configuration Topic	Service Monitor Configuration
Default Storage Location	`/var/lib/cloudera-service-monitor/` on the host where the Service Monitor role is configured to run.
Storage Configuration Defaults / Minimum / Maximum	• 10 GiB Services Time Series Storage • 1 GiB Impala Query Storage • 1 GiB YARN Application Storage Total: ~12 GiB Minimum (No Maximum)
Where to Control Data Retention or Size	Service Monitor data growth is controlled by configuring the maximum amount of storage space it can use.

Configuration Topic	Service Monitor Configuration
	To configure data retention in Cloudera Manager Administration Console: 1. Go the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Service Monitor or Cloudera Management Service (Service-Wide)**. 4. Select **Category** > **Main**. 5. Locate the ***propertyName*** property or search for it by typing its name in the Search box. **Time-Series Storage** The approximate amount of disk space dedicated to storing time series and health data. When the store has reached its maximum size, it deletes older data to make room for newer data. The disk usage is approximate because the store only begins deleting data when it reaches the limit. Note that Cloudera Manager stores time-series data at a number of different data granularities, and these granularities have different effective retention periods. The Service Monitor stores metric data not only as raw data points but also as ten-minute, hourly, six-hourly, daily, and weekly summary data points. Raw data consumes the bulk of the allocated storage space and weekly summaries consume the least. Raw data is retained for the shortest amount of time while weekly summary points are unlikely to ever be deleted. Select **Cloudera Management Service** > **Charts Library** tab in Cloudera Manager for information about how space is consumed within the Service Monitor. These pre-built charts also show information about the amount of data retained and time window covered by each data granularity. **Impala Storage** The approximate amount of disk space dedicated to storing Impala query data. When the store reaches its maximum size, it deletes older data to make room for newer queries. The disk usage is approximate because the store only begins deleting data when it reaches the limit. **YARN Storage** The approximate amount of disk space dedicated to storing YARN application data. When the store reaches its maximum size, it deletes older data to make room for newer applications. The disk usage is approximate because Cloudera Manager only begins deleting data when it reaches the limit. 6. Click **Save Changes** to commit the changes.
Sizing, Planning, and Best Practices	The Service Monitor gathers metrics about configured roles and services in your cluster and also runs active health tests. These health tests run regardless of idle and use periods, because they are always relevant. The Service Monitor gathers metrics and health test results regardless of the level of activity in the cluster. This data continues to grow, even in an idle cluster.

Table 4: Cloudera Management Service - Host Monitor

Configuration Topic	Host Monitor Configuration
Default Storage Location	`/var/lib/cloudera-host-monitor/` on the host where the Host Monitor role is configured to run.
Storage Configuration Defaults / Minimum/ Maximum	Default (and minimum): 10 GiB Host Time Series Storage
Where to Control Data Retention or Size	Host Monitor data growth is controlled by configuring the maximum amount of storage space it can use. See Data Storage for Monitoring Data. To configure these data retention configuration properties in the Cloudera Manager Administration Console: 1. Go the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Host Monitor or Cloudera Management Service (Service-Wide)**. 4. Select **Category** > **Main**. 5. Locate each property or search for it by typing its name in the Search box. **Time-Series Storage** The approximate amount of disk space dedicated to storing time series and health data. When the store reaches its maximum size, it deletes older data to make room for newer data. The disk usage is approximate because the store only begins deleting data when it reaches the limit. Note that Cloudera Manager stores time-series data at a number of different data granularities, and these granularities have different effective retention periods. Host Monitor stores metric data not only as raw data points but also as summaries of ten minute, one hour, six hour, one day, and one week increments. Raw data consumes the bulk of the allocated storage space and weekly summaries consume the least. Raw data is retained for the shortest amount of time, while weekly summary points are unlikely to ever be deleted. See the **Cloudera Management Service** > **Charts Library** tab in Cloudera Manager for information on how space is consumed within the Host Monitor. These pre-built charts also show information about the amount of data retained and the time window covered by each data granularity. 6. Click **Save Changes** to commit the changes.
Sizing, Planning and Best Practices	The Host Monitor gathers metrics about host-level items of interest (for example: disk space usage, RAM, CPU usage, swapping, etc) and also informs host health tests. The Host Monitor gathers metrics and health test results regardless of the level of activity in the cluster. This data continues to grow fairly linearly, even in an idle cluster.

Table 5: Cloudera Management Service - Event Server

Configuration Topic	Event Server Configuration
Default Storage Location	`/var/lib/cloudera-scm-eventserver/` on the host where the Event Server role is configured to run.

Configuration Topic	Event Server Configuration
Storage Configuration Defaults	5,000,000 events retained
Where to Control Data Retention or Minimum /Maximum	The amount of storage space the Event Server uses is influenced by configuring how many discrete events it can retain. To configure data retention in Cloudera Manager Administration Console, 1. Go the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Event Server or Cloudera Management Service (Service-Wide)**. 4. Select **Category** > **Main**. 5. Edit the following property: **Maximum Number of Events in the Event Server Store** The maximum size of the Event Server store, in events. When this size is exceeded, events are deleted starting with the oldest first until the size of the store is below this threshold 6. Click **Save Changes** to commit the changes.
Sizing, Planning, and Best Practices	The Event Server is a managed Lucene index that collects relevant events that happen within your cluster, such as results of health tests, log events that are created when a log entry matches a set of rules for identifying messages of interest and makes them available for searching, filtering and additional action. You can view and filter events on the **Diagnostics** > **Events** tab of the Cloudera Manager Administration Console. You can also poll this data using the Cloudera Manager API. **Note:** The Cloudera Management Service role Alert Publisher sources all the content for its work by regularly polling the Event Server for entries that are marked to be sent out using SNMP or SMTP(S). The Alert Publisher is not discussed because it has no noteworthy storage requirements of its own.

Table 6: Cloudera Management Service - Reports Manager

Configuration Topic	Reports Manager Configuration
Default Storage Location	**RDBMS:** Any Supported RDBMS. For more information, see CDH and Cloudera Manager Supported Databases. **Disk:** `/var/lib/cloudera-scm-headlamp/` on the host where the Reports Manager role is configured to run.
Storage Configuration Defaults	**RDBMS:** There are no configurable parameters to directly control the size of this data set. **Disk:** There are no configurable parameters to directly control the size of this data set. The storage utilization depends not only on the size of the HDFS *fsimage*,

Configuration Topic	Reports Manager Configuration
	but also on the HDFS file path complexity. Longer file paths contribute to more space utilization.
Where to Control Data Retention or Minimum / Maximum	The Reports Manager uses space in two main locations: on the Reports Manager host and on its supporting database. Cloudera recommends that the database be on a separate host from the Reports Manager host for process isolation and performance.
Sizing, Planning, and Best Practices	Reports Manager downloads the *fsimage* from the NameNode (every 60 minutes by default) and stores it locally to perform operations against, including indexing the HDFS filesystem structure. More files and directories results in a larger *fsimage*, which consumes more disk space. Reports Manager has no control over the size of the *fsimage*. If your total HDFS usage trends upward notably or you add excessively long paths in HDFS, it might be necessary to revisit and adjust the amount of local storage allocated to the Reports Manager. Periodically monitor, review, and adjust the local storage allocation.

Cloudera Navigator

Table 7: Cloudera Navigator - Navigator Audit Server

Configuration Topic	Navigator Audit Server Configuration
Default Storage Location	Any Supported RDBMS. For more information, see CDH and Cloudera Manager Supported Databases.
Storage Configuration Defaults	Default: 90 Days retention
Where to Control Data Retention or Min/Max	Navigator Audit Server storage usage is controlled by configuring how many days of data it can retain. Any older data is purged. To configure data retention in the Cloudera Manager Administration Console: 1. Go the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Navigator Audit Server or Cloudera Management Service (Service-Wide)**. 4. Select **Category** > **Main**. 5. Locate the **Navigator Audit Server Data Expiration Period** property or search for it by typing its name in the Search box. **Navigator Audit Server Data Expiration Period** In Navigator Audit Server, purge audit data of various auditable services when the data reaches this age in days. By default, Navigator Audit Server keeps data about audits for 90 days. 6. Click **Save Changes** to commit the changes.
Sizing, Planning, and Best Practices	The size of the Navigator Audit Server database directly depends on the number of audit events the cluster's audited services generate. Normally the volume of HDFS audits exceeds the volume of other audits (all other components like MRv1, Hive and Impala read from HDFS, which generates additional audit events). The average size of a discrete HDFS audit event is ~1 KB. For a busy cluster of 50 hosts with ~100K audit events generated per hour, the Navigator Audit

Configuration Topic	Navigator Audit Server Configuration
	Server database would consume ~2.5 GB per day. To retain 90 days of audits at that level, plan for a database size of around 250 GB. If other configured cluster services generate roughly the same amount of data as the HDFS audits, plan for the Navigator Audit Server database to require around 500 GB of storage for 90 days of data. Notes: • Individual Hive and Impala queries themselves can be very large. Since the query itself is part of an audit event, such audit events consume space in proportion to the length of the query. • The amount of space required increases as activity on the cluster increases. In some cases, Navigator Audit Server databases can exceed 1 TB for 90 days of audit events. Benchmark your cluster periodically and adjust accordingly. To map Cloudera Navigator versions to Cloudera Manager versions, see Product Compatibility Matrix for Cloudera Navigator.

Table 8: Cloudera Navigator - Navigator Metadata Server

Configuration Topic	Navigator Metadata Server Configuration
Default Storage Location	**RDBMS:** Any Supported RDBMS. For more information, see CDH and Cloudera Manager Supported Databases. **Disk:** `/var/lib/cloudera-scm-navigator/` on the host where the Navigator Metadata Server role is configured to run.
Storage Configuration Defaults	**RDBMS:** There are no exposed defaults or configurations to directly cull or purge the size of this data set. **Disk:** There are no configuration defaults to influence the size of this location. You can change the location itself with the **Navigator Metadata Server Storage Dir** property. The size of the data in this location depends on the amount of metadata in the system (HDFS fsimage size, Hive Metastore size) and activity on the system (the number of MapReduce Jobs run, Hive queries executed, etc).
Where to Control Data Retention or Min/Max	**RDBMS:** The Navigator Metadata Server database should be carefully tuned to support large volumes of metadata. **Disk:** The Navigator Metadata Server index (an embedded Solr instance) can consume lots of disk space at the location specified for the **Navigator Metadata Server Storage Dir** property. Ongoing maintenance tasks include purging metadata from the system.
Sizing, Planning, and Best Practices	**Memory:**

Configuration Topic	Navigator Metadata Server Configuration
	See Navigator Metadata Server Tuning. **RDBMS:** The database is used to store policies and authorization data. The dataset is small, but this database is also used during a Solr schema upgrade, where Solr documents are extracted and inserted again in Solr. This has same space requirements as above use case, but the space is only used temporarily during product upgrades. Use the Product Compatibility Matrix for Cloudera Navigator product compatibility matrix to map Cloudera Navigator and Cloudera Manager versions. **Disk:** This filesystem location contains all the metadata that is extracted from managed clusters. The data is stored in Solr, so this is the location where Solr stores its index and documents. Depending on the size of the cluster, this data can occupy tens of gigabytes. A guideline is to look at the size of HDFS fsimage and allocate two to three times that size as the initial size. The data here is incremental and continues to grow as activity is performed on the cluster. The rate of growth can be on order of tens of megabytes per day.

General Performance Notes

When possible:

- For entities that use an RDBMS, install the database on a separate host from the service, and consolidate roles that use databases on as few servers as possible.
- Provide a dedicated spindle to the RDBMS or datastore data directory to avoid disk contention with other read/write activity.

Cluster Lifecycle Management with Cloudera Manager

Cloudera Manager clusters that use parcels to provide CDH and other components require adequate disk space in the following locations:

Table 9: Parcel Lifecycle Management

Parcel Lifecycle Path (default)	Notes
Local Parcel Repository Path (`/opt/cloudera/parcel-repo`)	This path exists only on the host where Cloudera Manager Server (`cloudera-scm-server`) runs. The Cloudera Manager Server stages all new parcels in this location as it fetches them from any external repositories. Cloudera Manager Agents are then instructed to fetch the parcels from this location when the administrator distributes the parcel using the Cloudera Manager Administration Console or the Cloudera Manager API. **Sizing and Planning** The default location is `/opt/cloudera/parcel-repo` but you can configure another local filesystem location on the host where Cloudera Manager Server runs. See Parcel Configuration Settings. Provide sufficient space to hold all the parcels you download from all configured Remote Parcel Repository URLs (See Parcel Configuration Settings). Cloudera Manager deployments that manage multiple clusters store all applicable parcels for all clusters.

Parcel Lifecycle Path (default)	Notes
	Parcels are provided for each operating system, so be aware that heterogeneous clusters (distinct operating systems represented in the cluster) require more space than clusters with homogeneous operating systems. For example, a cluster with both RHEL6.x and 7.x hosts must hold -el6 and -el7 parcels in the Local Parcel Repository Path, which requires twice the amount of space. **Lifecycle Management and Best Practices** Delete any parcels that are no longer in use from the Cloudera Manager Administration Console, (never delete them manually from the command line) to recover disk space in the Local Parcel Repository Path and simultaneously across all managed cluster hosts which hold the parcel. **Backup Considerations** Perform regular backups of this path, and consider it a non-optional accessory to backing up Cloudera Manager Server. If you migrate Cloudera Manager Server to a new host or restore it from a backup (for example, after a hardware failure), recover the full content of this path to the new host, in the `/opt/cloudera/parcel-repo` directory before starting any `cloudera-scm-agent` or `cloudera-scm-server` processes.
Parcel Cache (`/opt/cloudera/parcel-cache`)	Managed Hosts running a Cloudera Manager Agent stage distributed parcels into this path (as `.parcel` files, unextracted). Do not manually manipulate this directory or its files. **Sizing and Planning** Provide sufficient space per-host to hold all the parcels you distribute to each host. You can configure Cloudera Manager to remove these cached `.parcel` files after they are extracted and placed in `/opt/cloudera/parcels/`. It is not mandatory to keep these temporary files but keeping them avoids the need to transfer the `.parcel` file from the Cloudera Manager Server repository should you need to extract the parcel again for any reason. To configure this behavior in the Cloudera Manager Administration Console, select **Administration** > **Settings** > **Parcels** > **Retain Downloaded Parcel Files**
Host Parcel Directory (`/opt/cloudera/parcels`)	Managed cluster hosts running a Cloudera Manager Agent extract parcels from the `/opt/cloudera/parcel-cache` directory into this path upon parcel activation. Many critical system symlinks point to files in this path and you should never manually manipulate its contents. **Sizing and Planning** Provide sufficient space on each host to hold all the parcels you distribute to each host. Be aware that the typical CDH parcel size is approximately 2 GB per parcel, and some third party parcels can exceed 3 GB. If you maintain various versions of parcels staged before and after upgrading, be aware of the disk space implications. You can configure Cloudera Manager to automatically remove older parcels when they are no longer in use. As an administrator you can always manually delete parcel versions not in use, but configuring these settings can handle the deletion automatically, in case you forget.

Parcel Lifecycle Path (default)	Notes
	To configure this behavior in the Cloudera Manager Administration Console, select **Administration** > **Settings** > **Parcels** and configure the following property: **Automatically Remove Old Parcels** This parameter controls whether parcels for old versions of an activated product should be removed from a cluster when they are no longer in use. The default value is Disabled. **Number of Old Parcel Versions to Retain** If you enable **Automatically Remove Old Parcels**, this setting specifies the number of old parcels to keep. Any old parcels beyond this value are removed. If this property is set to zero, no old parcels are retained. The default value is 3.

Table 10: Management Service Lifecycle - Space Reclamation Tasks

Task	Description
Activity Monitor (One-time)	The Activity Monitor only works against a MapReduce (MR1) service, not YARN. So if your deployment has fully migrated to YARN and no longer uses a MapReduce (MR1) service, your Activity Monitor database is no longer growing. If you have waited longer than the default Activity Monitor retention period (14 days) to address this point, then the Activity Monitor has already purged it all for you and your database is mostly empty. If your deployment meets these conditions, consider cleaning up by dropping the Activity Monitor database (again, only when you are satisfied that you no longer need the data or have confirmed that it is no longer in use) and the Activity Monitor role.
Service Monitor and Host Monitor (One-time)	For those who used Cloudera Manager version 4.x and have now upgraded to version 5.x: The Service Monitor and Host Monitor were migrated from their previously-configured RDBMS into a dedicated time series store used solely by each of these roles respectively. After this happens, there is still legacy database connection information in the configuration for these roles. This was used to allow for the initial migration but is no longer being used for any active work. After the above migration has taken place, the RDBMS databases previously used by the Service Monitor and Host Monitor are no longer used. Space occupied by these databases is now recoverable. If appropriate in your environment (and you are satisfied that you have long-term backups or do not need the data on disk any longer), you can drop those databases.
Ongoing Space Reclamation	Cloudera Management Services are automatically rolling up, purging or otherwise consolidating aged data for you in the background. Configure retention and purging limits per-role to control how and when this occurs. These configurations are discussed per-entity above. Adjust the default configurations to meet your space limitations or retention needs.

Log Files

All CDH cluster hosts write out separate log files for each role instance assigned to the host. Cluster administrators can monitor and manage the disk space used by these roles and configure log rotation to prevent log files from consuming too much disk space.

For more information, see Managing Disk Space for Log Files.

Conclusion

Keep this information in mind for planning and architecting the deployment of a cluster managed by Cloudera Manager. If you already have a live cluster, this lifecycle and backup information can help you keep critical monitoring, auditing, and metadata sources safe and properly backed up.

Configure Network Names

Important: CDH requires IPv4. IPv6 is not supported.

Tip: When bonding, use the `bond0` IP address as it represents all aggregated links.

Configure each host in the cluster as follows to ensure that all members can communicate with each other:

1. Set the hostname to a unique name (not `localhost`).

```
sudo hostnamectl set-hostname foo-1.example.com
```

2. Edit `/etc/hosts` with the IP address and fully qualified domain name (FQDN) of each host in the cluster. You can add the unqualified name as well.

```
1.1.1.1  foo-1.example.com  foo-1
2.2.2.2  foo-2.example.com  foo-2
3.3.3.3  foo-3.example.com  foo-3
4.4.4.4  foo-4.example.com  foo-4
```

Important:

- The canonical name of each host in `/etc/hosts` **must** be the FQDN (for example `myhost-1.example.com`), not the unqualified hostname (for example `myhost-1`). The canonical name is the first entry after the IP address.
- Do not use aliases, either in `/etc/hosts` or in configuring DNS.
- Unqualified hostnames (short names) must be unique in a Cloudera Manager instance. For example, you cannot have both *host01.example.com* and *host01.standby.example.com* managed by the same Cloudera Manager Server.

3. Edit `/etc/sysconfig/network` with the FQDN of this host only:

```
HOSTNAME=foo-1.example.com
```

4. Verify that each host consistently identifies to the network:
   a. Run `uname -a` and check that the hostname matches the output of the `hostname` command.
   b. Run `/sbin/ifconfig` and note the value of `inet addr` in the `eth0` (or `bond0`) entry, for example:

```
eth0      Link encap:Ethernet  HWaddr 00:0C:29:A4:E8:97
          inet addr:172.29.82.176  Bcast:172.29.87.255  Mask:255.255.248.0
...
```

   c. Run `host -v -t A $(hostname)` and verify that the output matches the `hostname` command.

      The IP address should be the same as reported by `ifconfig` for `eth0` (or `bond0`):

```
Trying "foo-1.example.com"
...
```

```
;; ANSWER SECTION:
foo-1.example.com. 60 IN A 172.29.82.176
```

Required Privileges for Package-based Installations of CDH

The following sections describe the user privilege requirements for package-based installation of CDH with Cloudera Manager. These requirements are standard UNIX system requirements for installing and managing packages and services.

Required Privileges

Important: Unless otherwise noted, when root or sudo access is required, using another system (such as PowerBroker) that provides root/sudo privileges is acceptable.

Table 11: Required Privileges for Package-Based CDH Installation

Task	Permissions Required
Install Cloudera Manager Server	`root` or `sudo` access to the host on which you are installing Cloudera Manager Server.
Start, stop, or restart Cloudera Manager Server using the `service` or `systemctl` utilities	`root` or `sudo` access to the Cloudera Manager Server host. The service runs as the `cloudera-scm` user by default.
Install CDH components using Cloudera Manager	One of the following, configured during initial installation of Cloudera Manager: • Access to the `root` user account using a password or SSH key file. • Passwordless `sudo` access for a specific user. For this task, using another system (such as PowerBroker) that provides `root` or `sudo` access is *not* supported.
Install Cloudera Manager Agent using Cloudera Manager	One of the following, configured during initial installation of Cloudera Manager: • Access to the `root` user account using a password or SSH key file. • Passwordless `sudo` access for a specific user. For this task, using another system (such as PowerBroker) that provides `root` or `sudo` access is *not* supported.
Automatically start Cloudera Manager Agent process	Access to the `root` user account during runtime, through one of the following scenarios: • During Cloudera Manager and CDH installation, the Agent is automatically started if installation is successful. It is then started using one of the following, as configured during the initial installation of Cloudera Manager: – Access to the `root` user account using a password or SSH key file. – Passwordless `sudo` access for a specific user. For this task, using another system (such as PowerBroker) that provides `root` or `sudo` access is *not* supported. • Through automatic startup during system boot, using `init`.

Task	Permissions Required
Manually start, stop, or restart Cloudera Manager Agent process	`root` or `sudo` access. This permission requirement ensures that services managed by the Cloudera Manager Agent can run as the appropriate user (such as the `hdfs` user for the HDFS service). Running commands within Cloudera Manager on a CDH service *does not* require root or `sudo` access, because the action is handled by the Cloudera Manager Agent, which is already running as the `root` user.

`sudo` Commands Run by Cloudera Manager

If you want to configure specific `sudo` access for the Cloudera Manager user (`cloudera-scm` by default), you can use the following list to do so.

The `sudo` commands run by Cloudera Manager are:

- `yum` (RHEL/CentOS/Oracle)
- `zypper` (SLES)
- `apt-get` (Ubuntu)
- `apt-key` (Ubuntu)
- `sed`
- `service`
- `/sbin/chkconfig` (RHEL/CentOS/Oracle)
- `/usr/sbin/update-rc.d` (Ubuntu)
- `id`
- `rm`
- `mv`
- `chown`
- `install`

Required Tomcat Directories

Most directories used by Cloudera services are configurable. Services that use Tomcat – for example, HttpFS, Oozie, Solr, Sqoop, KMS – require the directory `/var/lib/<service-specific-directory>`. For example, the Sqoop service requires the directory `/var/lib/sqoop`.

Ports

Cloudera Manager, CDH components, managed services, and third-party components use the ports listed in the tables that follow. Before you deploy Cloudera Manager, CDH, and managed services, and third-party components make sure these ports are open on each system. If you are using a firewall, such as `iptables` or `firewalld`, and cannot open all the listed ports, you must disable the firewall completely to ensure full functionality.

In the tables in the subsections that follow, the Access Requirement column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components (for example the JournalNode ports in an HA configuration); "External" means that the port can be used for either internal or external communication (for example, ports used by NodeManager and the JobHistory Server Web UIs).

Unless otherwise specified, the ports access requirement is unidirectional, meaning that inbound connections to the specified ports must be allowed. In most modern stateful firewalls, it is not necessary to create a separate rule for return traffic on a permitted session.

Ports Used by Cloudera Manager and Cloudera Navigator

The following diagram provides an overview of some of the ports used by Cloudera Manager, Cloudera Navigator, and Cloudera Management Service roles:

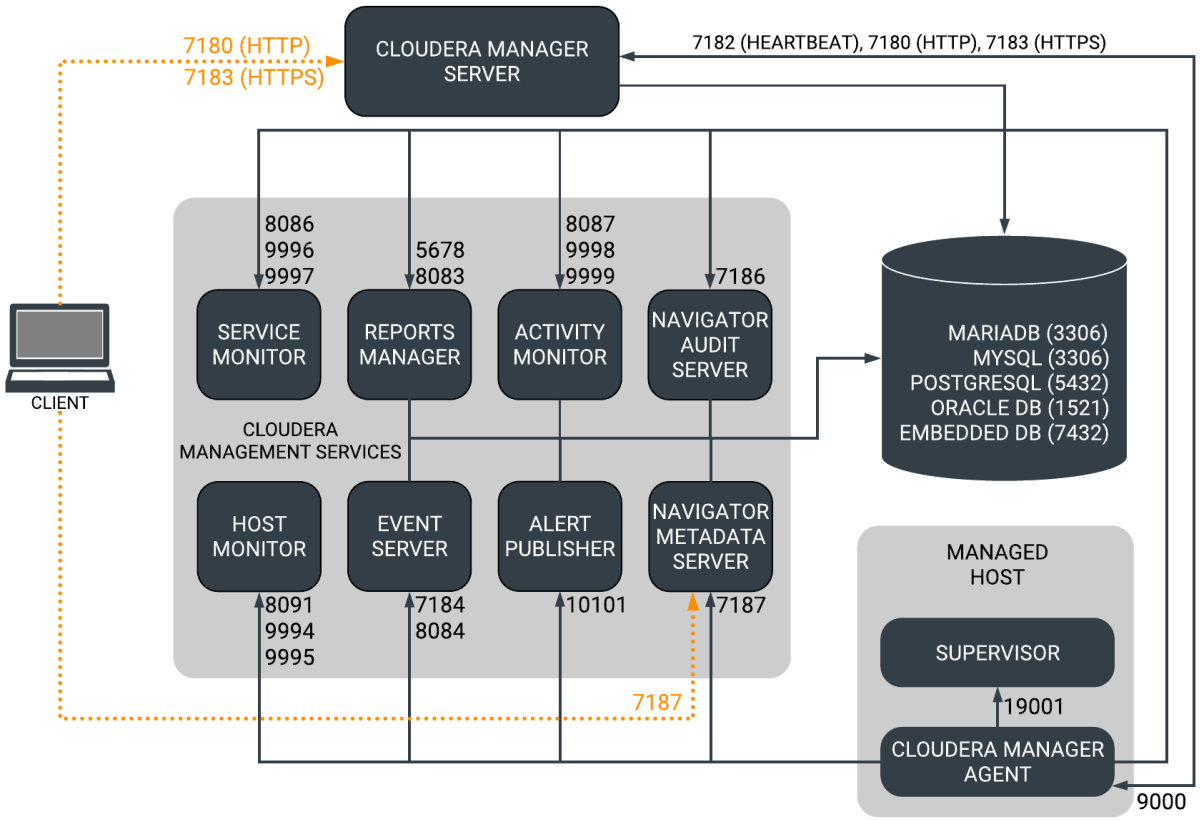


Figure 1: Ports Used by Cloudera Manager and Cloudera Navigator

When peer-to-peer distribution is enabled for parcels, the Cloudera Manager Agent can obtain the parcel from the Cloudera Manager Server or from other agents, as follows:

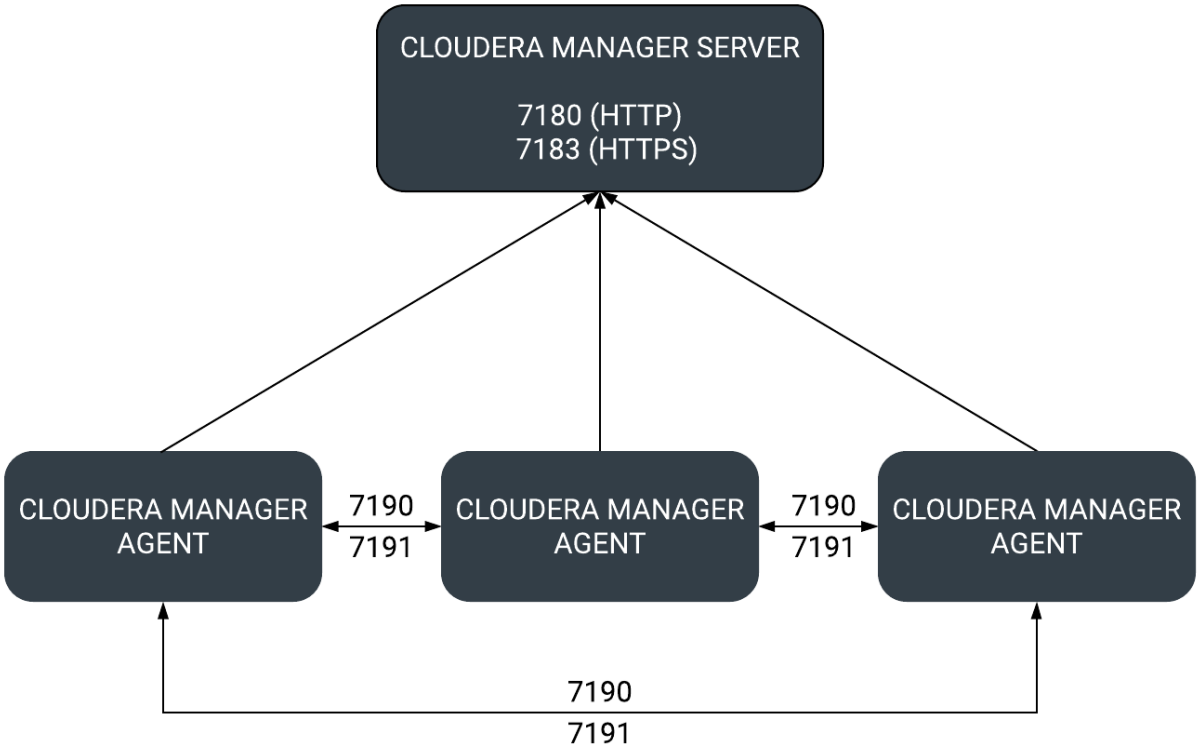


Figure 2: Ports Used in Peer-to-Peer Parcel Distribution

For further details, see the following tables. All ports listed are TCP.

In the following tables, *Internal* means that the port is used only for communication among the components; *External* means that the port can be used for either internal or external communication.

Table 12: External Ports

Component	Service	Port	Configuration	Description
Cloudera Manager Server	HTTP (Web UI)	7180	**Administration** > **Settings** > **Category** > **Ports and Addresses** > **HTTP Port for Admin Console**	HTTP port used by the web console.
	HTTPS (Web UI)	7183	**Administration** > **Settings** > **Category** > **Ports and Addresses** > **HTTPS Port for Admin Console**	Port used by the web console if HTTPS is enabled. If enabled, port 7180 remains open, but redirects all requests to HTTPS on port 7183.
Cloudera Navigator Metadata Server	HTTP (Web UI)	7187	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Navigator Metadata Server Port**	The port where Navigator Metadata Server listens for requests.

Component	Service	Port	Configuration	Description
Backup and Disaster Recovery	HTTP (Web UI)	7180	**Administration** > **Settings** > **Category** > **Ports and Addresses** > **HTTP Port for Admin Console**	Used for communication to peer (source) Cloudera Manager.
	HTTPS (Web UI)	7183	**Administration** > **Settings** > **Category** > **Ports and Addresses** > **HTTPS Port for Admin Console**	Used for communication to peer (source) Cloudera Manager when HTTPS is enabled.
	HDFS NameNode	8020	**HDFS service** > **Configuration** > **Category** > **Ports and Addresses** > **NameNode Port**	HDFS and Hive/Impala replication: communication from destination HDFS and MapReduce hosts to source HDFS NameNode(s). Hive/Impala Replication: communication from source Hive hosts to destination HDFS NameNode(s).
	HDFS DataNode	50010	**HDFS service** > **Configuration** > **Category** > **Ports and Addresses** > **DataNode Transceiver Port**	HDFS and Hive/Impala replication: communication from destination HDFS and MapReduce hosts to source HDFS DataNode(s). Hive/Impala Replication: communication from source Hive hosts to destination HDFS DataNode(s).
Telemetry Publisher	HTTP	10110	**Clusters** > **Cloudera Management Service** > **Category** > **Ports and Addresses** > **Telemetry Publisher Server Port**	The port where the Telemetry Publisher Server listens for requests
Telemetry Publisher	HTTP (Debug)	10111	**Clusters** > **Cloudera Management Service** > **Category** > **Ports and Addresses** > **Telemetry Publisher Web UI Port**	The port where Telemetry Publisher starts a debug web server. Set to -1 to disable debug server.

Table 13: Internal Ports

Component	Service	Port	Configuration	Description
Cloudera Manager Server	Avro (RPC)	7182	**Administration** > **Settings** > **Category** > **Ports and Addresses** > **Agent Port to connect to Server**	Used for Agent to Server heartbeats
	Embedded PostgreSQL database	7432		The optional embedded PostgreSQL database used for storing configuration information for Cloudera Manager Server.
	Peer-to-peer parcel distribution	7190, 7191	**Hosts** > **All Hosts** > **Configuration** > **P2P Parcel Distribution Port**	Used to distribute parcels to cluster hosts during installation and upgrade operations.
Cloudera Manager Agent	HTTP (Debug)	9000	/etc/cloudera-scm-agent/config.ini	
Event Server	Custom protocol	7184	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Event Publish Port**	Port on which the Event Server listens for the publication of events.
	Custom protocol	7185	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Event Query Port**	Port on which the Event Server listens for queries for events.
	HTTP (Debug)	8084	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Event Server Web UI Port**	Port for the Event Server's Debug page. Set to -1 to disable debug server.
Alert Publisher	Custom protocol	10101	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Alerts: Listen Port**	Port where the Alert Publisher listens for internal API requests.
Service Monitor	HTTP (Debug)	8086	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Service Monitor Web UI Port**	Port for Service Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)		**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Service Monitor Web UI HTTPS Port**	Port for Service Monitor's HTTPS Debug page.
	Custom protocol	9997	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Service Monitor Listen Port**	Port where Service Monitor is listening for agent messages.

Component	Service	Port	Configuration	Description
	Internal query API (Avro)	9996	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Service Monitor Nozzle Port**	Port where Service Monitor's query API is exposed.
Activity Monitor	HTTP (Debug)	8087	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Activity Monitor Web UI Port**	Port for Activity Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)		**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Activity Monitor Web UI HTTPS Port**	Port for Activity Monitor's HTTPS Debug page.
	Custom protocol	9999	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Activity Monitor Listen Port**	Port where Activity Monitor is listening for agent messages.
	Internal query API (Avro)	9998	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Activity Monitor Nozzle Port**	Port where Activity Monitor's query API is exposed.
Host Monitor	HTTP (Debug)	8091	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Host Monitor Web UI Port**	Port for Host Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)	9091	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Host Monitor Web UI HTTPS Port**	Port for Host Monitor's HTTPS Debug page.
	Custom protocol	9995	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Host Monitor Listen Port**	Port where Host Monitor is listening for agent messages.
	Internal query API (Avro)	9994	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Host Monitor Nozzle Port**	Port where Host Monitor's query API is exposed.
Reports Manager	Queries (Thrift)	5678	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and**	The port where Reports Manager listens for requests.

Component	Service	Port	Configuration	Description
			Addresses > **Reports Manager Server Port**	
	HTTP (Debug)	8083	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Reports Manager Web UI Port**	The port where Reports Manager starts a debug web server. Set to -1 to disable debug server.
Cloudera Navigator Audit Server	HTTP	7186	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Navigator Audit Server Port**	The port where Navigator Audit Server listens for requests.
	HTTP (Debug)	8089	**Cloudera Management Service** > **Configuration** > **Category** > **Ports and Addresses** > **Navigator Audit Server Web UI Port**	The port where Navigator Audit Server runs a debug web server. Set to -1 to disable debug server.

Ports Used by Cloudera Navigator Encryption

All ports listed are TCP.

In the following table, the **Access Requirement** column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components; "External" means that the port can be used for either internal or external communication.

Component	Service	Port	Access Requirement	Configuration	Comment
Cloudera Navigator Key Trustee Server	HTTPS (key management)	11371	External	**Key Trustee Server service** > **Configuration** > **Category** > **Ports and Addresses** > **Key Trustee Server Port**	Navigator Key Trustee Server clients (including Key Trustee KMS and Navigator Encrypt) access this port to store and retrieve encryption keys.
	PostgreSQL database	11381	External	**Key Trustee Server service** > **Configuration** > **Category** > **Ports and Addresses** > **Key Trustee Server Database Port**	The Navigator Key Trustee Server database listens on this port. The Passive Key Trustee Server connects to this port on the Active Key Trustee Server for replication in Cloudera Navigator Key Trustee Server High Availability.

Ports Used by Components of CDH

All ports listed are TCP.

In the following table, the **Access Requirement** column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components; "External" means that the port can be used for either internal or external communication.

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
Hadoop HDFS	DataNode		50010	External	`dfs.datanode.address`	DataNode HTTP server port
	DataNode	Secure	1004	External	`dfs.datanode.address`	
	DataNode		50075	External	`dfs.datanode.http.address`	
	DataNode		50475	External	`dfs.datanode.https.address`	
	DataNode	Secure	1006	External	`dfs.datanode.http.address`	
	DataNode		50020	External	`dfs.datanode.ipc.address`	
	NameNode		8020	External	`fs.default.name` or `fs.defaultFS`	`fs.default.name` is deprecated (but still works)
	NameNode		8022	External	`dfs.namenode.servicerpc-address`	Optional port used by HDFS daemons to avoid sharing the RPC port used by clients (8020). Cloudera recommends using port 8022.
	NameNode		50070	External	`dfs.http.address` or `dfs.namenode.http-address`	`dfs.http.address` is deprecated (but still works)
	NameNode	Secure	50470	External	`dfs.https.address` or `dfs.namenode.https-address`	`dfs.https.address` is deprecated (but still works)
	Secondary NameNode		50090	Internal	`dfs.secondary.http.address` or `dfs.namenode.secondary.http-address`	`dfs.secondary.http.address` is deprecated (but still works)
	Secondary NameNode	Secure	50495	Internal	`dfs.secondary.https.address`	
	JournalNode		8485	Internal	`dfs.namenode.shared.edits.dir`	
	JournalNode		8480	Internal	`dfs.journalnode.http-address`	
	JournalNode		8481	Internal	`dfs.journalnode.https-address`	
	Failover Controller		8019	Internal		Used for NameNode HA
	NFS gateway		2049	External		`nfs` port (`nfs3.server.port`)

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
	NFS gateway		4242	External		`mountd` port (`nfs3.mountd.port`)
	NFS gateway		111	External		`portmapper` or `rpcbind` port
	NFS gateway		50079	External	`nfs.http.port`	CDH 5.4.0 and higher. The NFS gateway daemon uses this port to serve metrics. The port is configurable on versions 5.10 and higher.
	NFS gateway	Secure	50579	External	`nfs.https.port`	CDH 5.4.0 and higher. The NFS gateway daemon uses this port to serve metrics. The port is configurable on versions 5.10 and higher.
	HttpFS		14000	External		
	HttpFS		14001	External		
Hadoop YARN (MRv2)	ResourceManager		8032	External	`yarn.resourcemanager.address`	
	ResourceManager		8030	Internal	`yarn.resourcemanager.scheduler.address`	
	ResourceManager		8031	Internal	`yarn.resourcemanager.resource-tracker.address`	
	ResourceManager		8033	External	`yarn.resourcemanager.admin.address`	
	ResourceManager		8088	External	`yarn.resourcemanager.webapp.address`	
	ResourceManager		8090	External	`yarn.resourcemanager.webapp.https.address`	
	NodeManager		8040	Internal	`yarn.nodemanager.localizer.address`	
	NodeManager		8041	Internal	`yarn.nodemanager.address`	
	NodeManager		8042	External	`yarn.nodemanager.webapp.address`	

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
	NodeManager		8044	External	`yarn.nodemanager.webapp.https.address`	
	JobHistory Server		10020	Internal	`mapreduce.jobhistory.address`	
	JobHistory Server		10033	Internal	`mapreduce.jobhistory.admin.address`	
	Shuffle HTTP		13562	Internal	`mapreduce.shuffle.port`	
	JobHistory Server		19888	External	`mapreduce.jobhistory.webapp.address`	
	JobHistory Server		19890	External	`mapreduce.jobhistory.webapp.https.address`	
	ApplicationMaster			External		The ApplicationMaster serves an HTTP service using an ephemeral port that cannot be restricted. This port is never accessed directly from outside the cluster by clients. All requests to the ApplicationMaster web server is routed using the YARN ResourceManager (proxy service). Locking down access to ephemeral port ranges within the cluster's network might restrict your access to the ApplicationMaster UI and its logs, along with the ability to look at running applications.
Flume	Flume Agent		41414	External		
Hadoop KMS	Key Management Server		16000	External	`kms_http_port`	CDH 5.2.1 and higher. Applies to both Java KeyStore KMS and Key Trustee KMS.
	Key Management Server		16001	Localhost	`kms_admin_port`	CDH 5.2.1 and higher. Applies to both Java KeyStore KMS and Key Trustee KMS.
HBase	Master		60000	External	`hbase.master.port`	IPC
	Master		60010	External	`hbase.master.info.port`	HTTP

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
	RegionServer		60020	External	`hbase.regionserver.port`	IPC
	RegionServer		60030	External	`hbase.regionserver.info.port`	HTTP
	HQuorumPeer		2181	Internal	`hbase.zookeeper.property.clientPort`	HBase-managed ZooKeeper mode
	HQuorumPeer		2888	Internal	`hbase.zookeeper.peerport`	HBase-managed ZooKeeper mode
	HQuorumPeer		3888	Internal	`hbase.zookeeper.leaderport`	HBase-managed ZooKeeper mode
	REST	Non-Cloudera Manager-managed	8080	External	`hbase.rest.port`	The default REST port in HBase is 8080. Because this is a commonly used port, Cloudera Manager sets the default to 20550 instead.
	REST	Cloudera Manager-managed	20550	External	`hbase.rest.port`	The default REST port in HBase is 8080. Because this is a commonly used port, Cloudera Manager sets the default to 20550 instead.
	REST UI		8085	External		
	Thrift Server	Thrift Server	9090	External	Pass `-p <port>` on CLI	
	Thrift Server		9095	External		
		Avro server	9090	External	Pass `--port <port>` on CLI	
	hbase-solr-indexer	Lily Indexer	11060	External		
Hive	Metastore		9083	External		
	HiveServer2		10000	External	`hive.server2.thrift.port`	The Beeline command interpreter requires that you specify this port on the command line. If you use Oracle database, you must manually reserve this port. For more information, see Reserving Ports for HiveServer 2 on page 105.

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
	HiveServer2 Web User Interface (UI)		10002	External	`hive.server2.webui.port` in `hive-site.xml`	
	WebHCat Server		50111	External	`templeton.port`	
Hue	Server		8888	External		
	Load Balancer		8889	External		
Kafka	Broker	TCP Port	9092	External/Internal	port	The primary communication port used by producers and consumers; also used for inter-broker communication.
	Broker	TLS/SSL Port	9093	External/Internal	ssl_port	A secured communication port used by producers and consumers; also used for inter-broker communication.
	Broker	JMX Port	9393	Internal	jmx_port	Internal use only. Used for administration via JMX.
	MirrorMaker	JMX Port	9394	Internal	jmx_port	Internal use only. Used to administer the producer and consumer of the MirrorMaker.
	Broker	HTTP Metric Report Port	24042	Internal	kafka.http.metrics.port	Internal use only. This is the port via which the HTTP metric reporter listens. It is used to retrieve metrics through HTTP instead of JMX.
Kudu	Master		7051	External		Kudu Master RPC port
	Master		8051	External		Kudu Master HTTP server port
	TabletServer		7050	External		Kudu TabletServer RPC port
	TabletServer		8050	External		Kudu TabletServer HTTP server port
Oozie	Oozie Server		11000	External	`OOZIE_HTTP_PORT` in `oozie-env.sh`	HTTP
	Oozie Server	SSL	11443	External		HTTPS

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
Sentry	Sentry Server		8038	External	`sentry.service.server.rpc-port`	
	Sentry Server		51000	External	`sentry.service.web.port`	
Solr	Solr Server		8983	External		HTTP port for all Solr-specific actions, update/query.
	Solr Server		8984	Internal		Solr administrative use.
	Solr Server		8985	External		HTTPS port for all Solr-specific actions, update/query.
Spark	Default Master RPC port		7077	External		
	Default Worker RPC port		7078	External		
	Default Master web UI port		18080	External		
	Default Worker web UI port		18081	External		
	History Server		18088	External	`history.port`	
	Shuffle service		7337	Internal		
Sqoop	Metastore		16000	External	`sqoop.metastore.server.port`	
Sqoop 2	Sqoop 2 server		8005	Localhost	`SQOOP_ADMIN_PORT` environment variable	
	Sqoop 2 server		12000	External		
	Sqoop 2		12001	External		Admin port
ZooKeeper	Server (with CDH 5 or Cloudera Manager 5)		2181	External	clientPort	Client port
	Server (with CDH 5 only)		2888	Internal	`X in server.N=host:X:Y`	Peer
	Server (with CDH 5 only)		3888	Internal	`X in server.N=host:X:Y`	Peer
	Server (with CDH 5 and Cloudera Manager 5)		3181	Internal	`X in server.N=host:X:Y`	Peer

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
	Server (with CDH 5 and Cloudera Manager 5)		4181	Internal	`X in server.N =host:X:Y`	Peer
	ZooKeeper JMX port		9010	Internal		ZooKeeper will also use another randomly selected port for RMI. To allow Cloudera Manager to monitor ZooKeeper, you must do *one* of the following: • Open up all ports when the connection originates from the Cloudera Manager Server • Do the following: 1. Open a non-ephemeral port (such as 9011) in the firewall. 2. Install Oracle Java 7u4 JDK or higher. 3. Add the port configuration to the advanced configuration snippet, for example: `-Dcom.sun.management.jmxremote.rmi.port=9011` 4. Restart ZooKeeper.

Ports Used by Impala

Impala uses the TCP ports listed in the following table. Before deploying Impala, ensure these ports are open on each system.

Scope / Role	Cloudera Manager Setting	Default Port	Comment
Impala Daemon	Impala Daemon Frontend Port	21000	Used to transmit commands and receive results by `impala-shell` and version 1.2 of the Cloudera ODBC driver.
Impala Daemon	Impala Daemon Frontend Port	21050	Used to transmit commands and receive results by applications, such

Scope / Role	Cloudera Manager Setting	Default Port	Comment
			as Business Intelligence tools, using JDBC, the Beeswax query editor in Hue, and version 2.0 or higher of the Cloudera ODBC driver.
Impala Daemon	Impala Daemon Backend Port	22000	Internal use only. Impala daemons use this port to communicate with each other.
Impala Daemon	StateStoreSubscriber Service Port	23000	Internal use only. Impala daemons listen on this port for updates from the statestore daemon.
Catalog Daemon	StateStoreSubscriber Service Port	23020	Internal use only. The catalog daemon listens on this port for updates from the statestore daemon.
Impala Daemon	Impala Daemon HTTP Server Port	25000	Impala web interface for administrators to monitor and troubleshoot.
Impala StateStore Daemon	StateStore HTTP Server Port	25010	StateStore web interface for administrators to monitor and troubleshoot.
Impala Catalog Daemon	Catalog HTTP Server Port	25020	Catalog service web interface for administrators to monitor and troubleshoot. New in Impala 1.2 and higher.
Impala StateStore Daemon	StateStore Service Port	24000	Internal use only. The statestore daemon listens on this port for registration/unregistration requests.
Impala Catalog Daemon	Catalog Service Port	26000	Internal use only. The catalog service uses this port to communicate with the Impala daemons. New in Impala 1.2 and higher.
Impala Daemon	KRPC Port	27000	Internal use only. The port on which KRPC-based Impala Internal Service is exported. You can change the default value of 27000 with the `--krpc_port` startup flag.
Impala Daemon	Llama Callback Port	28000	Internal use only. Impala daemons use to communicate with Llama. New in and higher.
	`krpc_port`	27000	
Impala Llama ApplicationMaster	Llama Thrift Admin Port	15002	Internal use only. New in and higher.

Scope / Role	Cloudera Manager Setting	Default Port	Comment
Impala Llama ApplicationMaster	Llama Thrift Port	15000	Internal use only. New in and higher.
Impala Llama ApplicationMaster	Llama HTTP Port	15001	Llama service web interface for administrators to monitor and troubleshoot. New in and higher.

Ports Used by Cloudera Search

In the following table, the **Access Requirement** column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components; "External" means that the port can be used for either internal or external communication.

Component	Service	Port	Protocol	Access Requirement	Comment
Cloudera Search	Solr search/update	8983	http	External	All Solr-specific actions, update/query
Cloudera Search	Solr (admin)	8984	http	Internal	Solr administrative use
Cloudera Search	Solr (server)	8985	https	External	All Solr-specific actions, update/query

Ports Used by DistCp

All ports listed are TCP.

In the following table, the **Access Requirement** column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components; "External" means that the port can be used for either internal or external communication.

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment

Hadoop HDFS	NameNode		8020	External	`fs.default.name`
					or
					`fs.defaultFS`
						`fs.default.name`
						is deprecated (but still works)

	DataNode	Secure	1004	External	`dfs.datanode.address`	
	DataNode		50010	External	`dfs.datanode.address`	

WebHDFS	NameNode		50070	External	`dfs.http.address`
					or
					`dfs.namenode.http-address`
						`dfs.http.address`
						is deprecated (but still works)

	DataNode	Secure	1006	External	`dfs.datanode.http.address`	

Component	Service	Qualifier	Port	Access Requirement	Configuration	Comment
HttpFS	web		14000			

Ports Used by Third-Party Components

In the following table, the **Access Requirement** column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components; "External" means that the port can be used for either internal or external communication.

Component	Service	Qualifier	Port	Protocol	Access Requirement	Configuration	Comment
Ganglia	ganglia-gmond		8649	UDP/TCP	Internal		
	ganglia-web		80	TCP	External	Via Apache `httpd`	
Kerberos	KRB5 KDC Server	Secure	88	UDP/TCP	External	`kdc_ports` and `kdc_tcp_ports` in either the `[kdcdefaults]` or `[realms]` sections of `kdc.conf`	By default only UDP
	KRB5 Admin Server	Secure	749	TCP	External	`kadmind_port` in the `[realms]` section of `kdc.conf`	
	kpasswd		464	UDP/TCP	External		
SSH	ssh		22	TCP	External		
PostgreSQL			5432	TCP	Internal		
MariaDB			3306	TCP	Internal		
MySQL			3306	TCP	Internal		
LDAP	LDAP Server		389	TCP	External		
	LDAP Server over TLS/SSL	TLS/SSL	636	TCP	External		
	Global Catalog		3268	TCP	External		
	Global Catalog over TLS/SSL	TLS/SSL	3269	TCP	External		

Ports Used by Apache Flume and Apache Solr

Ports used by Apache Flume to communicate with Apache Solr might vary, depending on your configuration and whether you are using security (for example, SSL). A typical cluster set up with Flume writing to HDFS and Solr sinks uses the ports listed below:

Port	Description
41414	HTTP Port - port on which the Flume web server listens for requests. Flume uses this port continuously.
7184	Cloudera Manager Event Publish Port - port on which the Event Server listens for the publication of events. Flume uses this port continuously.
8020	NameNode Port, used by the HDFS sink.
8983	Solr HTTP Port, used by MorphlineSolrSink.
2181	ZooKeeper client port, used by MorphlineSolrSink.

Recommended Cluster Hosts and Role Distribution

Important: This topic describes suggested role assignments for a CDH cluster managed by Cloudera Manager. The actual assignments you choose for your deployment can vary depending on the types and volume of work loads, the services deployed in your cluster, hardware resources, configuration, and other factors.

When you install CDH using the Cloudera Manager installation wizard, Cloudera Manager attempts to spread the roles among cluster hosts (except for roles assigned to gateway hosts) based on the resources available in the hosts. You can change these assignments on the **Customize Role Assignments** page that appears in the wizard. You can also change and add roles at a later time using Cloudera Manager. See Role Instances.

If your cluster uses data-at-rest encryption, see Allocating Hosts for Key Trustee Server and Key Trustee KMS on page 43.

For information about where to locate various databases that are required for Cloudera Manager and other services, see Step 4: Install and Configure Databases on page 87.

CDH Cluster Hosts and Role Assignments

Cluster hosts can be broadly described as the following types:

- **Master hosts** run Hadoop master processes such as the HDFS NameNode and YARN Resource Manager.
- **Utility hosts** run other cluster processes that are not master processes such as Cloudera Manager and the Hive Metastore.
- **Gateway hosts** are client access points for launching jobs in the cluster. The number of gateway hosts required varies depending on the type and size of the workloads.
- **Worker hosts** primarily run DataNodes and other distributed processes such as Impalad.

Important: Cloudera recommends that you always enable high availability when CDH is used in a production environment.

The following tables describe the recommended role allocations for different cluster sizes:

3 - 10 Worker Hosts without High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode • YARN ResourceManager • JobHistory Server • ZooKeeper • Kudu master • Spark History Server	**One host for all Utility and Gateway roles:** • Secondary NameNode • Cloudera Manager • Cloudera Manager Management Service • Hive Metastore • HiveServer2 • Impala Catalog Server • Impala StateStore • Hue • Oozie • Flume • Gateway configuration		**3 - 10 Worker Hosts:** • DataNode • NodeManager • Impalad • Kudu tablet server

3 - 20 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode • JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • JobHistory Server • Spark History Server • Kudu master **Master Host 2:** • NameNode • JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • Kudu master **Master Host 3:** • Kudu master (Kudu requires an odd number of masters for HA.)	**Utility Host 1:** • Cloudera Manager • Cloudera Manager Management Service • Hive Metastore • Impala Catalog Server • Impala StateStore • Oozie • ZooKeeper (requires dedicated disk) • JournalNode (requires dedicated disk)	**One or more Gateway Hosts:** • Hue • HiveServer2 • Flume • Gateway configuration	**3 - 20 Worker Hosts:** • DataNode • NodeManager • Impalad • Kudu tablet server

20 - 80 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode	**Utility Host 1:** • Cloudera Manager	**One or more Gateway Hosts:** • Hue	**20 - 80 Worker Hosts:** • DataNode

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
• JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • Kudu master **Master Host 2:** • NameNode • JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • Kudu master **Master Host 3:** • ZooKeeper • JournalNode • JobHistory Server • Spark History Server • Kudu master	**Utility Host 2:** • Cloudera Manager Management Service • Hive Metastore • Impala Catalog Server • Oozie	• HiveServer2 • Flume • Gateway configuration	• NodeManager • Impalad • Kudu tablet server

80 - 200 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode • JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • Kudu master **Master Host 2:** • NameNode • JournalNode • FailoverController • YARN ResourceManager • ZooKeeper • Kudu master **Master Host 3:** • ZooKeeper • JournalNode • JobHistory Server • Spark History Server • Kudu master	**Utility Host 1:** • Cloudera Manager **Utility Host 2:** • Hive Metastore • Impala Catalog Server • Impala StateStore • Oozie **Utility Host 3:** • Activity Monitor **Utility Host 4:** • Host Monitor **Utility Host 5:** • Navigator Audit Server **Utility Host 6:** • Navigator Metadata Server **Utility Host 7:** • Reports Manager	**One or more Gateway Hosts:** • Hue • HiveServer2 • Flume • Gateway configuration	**80 - 200 Worker Hosts:** • DataNode • NodeManager • Impalad • Kudu tablet server (Recommended maximum number of tablet servers is 100.)

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
	Utility Host 8: • Service Monitor		

200 - 500 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode • JournalNode • FailoverController • ZooKeeper • Kudu master **Master Host 2:** • NameNode • JournalNode • FailoverController • ZooKeeper • Kudu master **Master Host 3:** • YARN ResourceManager • ZooKeeper • JournalNode • Kudu master **Master Host 4:** • YARN ResourceManager • ZooKeeper • JournalNode **Master Host 5:** • JobHistory Server • Spark History Server • ZooKeeper • JournalNode We recommend no more than three Kudu masters.	**Utility Host 1:** • Cloudera Manager **Utility Host 2:** • Hive Metastore • Impala Catalog Server • Impala StateStore • Oozie **Utility Host 3:** • Activity Monitor **Utility Host 4:** • Host Monitor **Utility Host 5:** • Navigator Audit Server **Utility Host 6:** • Navigator Metadata Server **Utility Host 7:** • Reports Manager **Utility Host 8:** • Service Monitor	**One or more Gateway Hosts:** • Hue • HiveServer2 • Flume • Gateway configuration	**200 - 500 Worker Hosts:** • DataNode • NodeManager • Impalad • Kudu tablet server (Recommended maximum number of tablet servers is 100.)

500 -1000 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: • NameNode • JournalNode • FailoverController • ZooKeeper	**Utility Host 1:** • Cloudera Manager **Utility Host 2:** • Hive Metastore	**One or more Gateway Hosts:** • Hue • HiveServer2 • Flume	**500 - 1000 Worker Hosts:** • DataNode • NodeManager • Impalad • Kudu tablet server (Recommended

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
• Kudu master **Master Host 2:** • NameNode • JournalNode • FailoverController • ZooKeeper • Kudu master **Master Host 3:** • YARN ResourceManager • ZooKeeper • JournalNode • Kudu master **Master Host 4:** • YARN ResourceManager • ZooKeeper • JournalNode **Master Host 5:** • JobHistory Server • Spark History Server • ZooKeeper • JournalNode We recommend no more than three Kudu masters.	• Impala Catalog Server • Impala StateStore • Oozie **Utility Host 3:** • Activity Monitor **Utility Host 4:** • Host Monitor **Utility Host 5:** • Navigator Audit Server **Utility Host 6:** • Navigator Metadata Server **Utility Host 7:** • Reports Manager **Utility Host 8:** • Service Monitor	• Gateway configuration	maximum number of tablet servers is 100.)

Allocating Hosts for Key Trustee Server and Key Trustee KMS

If you are enabling data-at-rest encryption for a CDH cluster, Cloudera recommends that you isolate the Key Trustee Server from other enterprise data hub (EDH) services by deploying the Key Trustee Server on dedicated hosts in a separate cluster managed by Cloudera Manager. Cloudera also recommends deploying Key Trustee KMS on dedicated hosts in the same cluster as the EDH services that require access to Key Trustee Server. This architecture helps users avoid having to restart the Key Trustee Server when restarting a cluster.

For more information about encrypting data at rest in an EDH, see Encrypting Data at Rest.

For production environments in general, or if you have enabled high availability for HDFS and are using data-at-rest encryption, Cloudera recommends that you enable high availability for Key Trustee Server and Key Trustee KMS.

See:

- Cloudera Navigator Key Trustee Server High Availability
- Enabling Key Trustee KMS High Availability

Custom Installation Solutions

Cloudera hosts two types of software repositories that you can use to install products such as Cloudera Manager or CDH—parcel repositories and package repositories.

These repositories are effective solutions in most cases, but custom installation solutions are sometimes required. Using the Cloudera-hosted software repositories requires client access over the Internet. Typical installations use the latest available software. In some scenarios, these behaviors might not be desirable, such as:

- You need to install older product versions. For example, in a CDH cluster, all hosts must run the same CDH version. After completing an initial installation, you may want to add hosts. This could be to increase the size of your cluster to handle larger tasks or to replace older hardware.
- The hosts on which you want to install Cloudera products are not connected to the Internet, so they cannot reach the Cloudera repository. (For a parcel installation, only the Cloudera Manager Server needs Internet access, but for a package installation, all cluster hosts require access to the Cloudera repository). Most organizations partition parts of their network from outside access. Isolating network segments improves security, but can add complexity to the installation process.

In both of these cases, using an internal repository allows you to meet the needs of your organization, whether that means installing specific versions of Cloudera software or installing Cloudera software on hosts without Internet access.

Introduction to Parcels

Parcels are a packaging format that facilitate upgrading software from within Cloudera Manager. You can download, distribute, and activate a new software version all from within Cloudera Manager. Cloudera Manager downloads a parcel to a local directory. Once the parcel is downloaded to the Cloudera Manager Server host, an Internet connection is no longer needed to deploy the parcel. For detailed information about parcels, see Parcels.

If your Cloudera Manager Server does not have Internet access, you can obtain the required parcel files and put them into a parcel repository. For more information, see Using an Internal Parcel Repository on page 45.

Understanding Package Management

Before getting into the details of how to configure a custom package management solution in your environment, it can be useful to have more information about:

Package Management Tools

Packages (`rpm` or deb files) help ensure that installations complete successfully by satisfying package dependencies. When you install a particular package, all other required packages are installed at the same time. For example, `hadoop-0.20-hive` depends on `hadoop-0.20`.

Package management tools, such as `yum` (RHEL), `zypper` (SLES), and `apt-get` (Ubuntu) are tools that can find and install required packages. For example, on a RHEL compatible system, you might run the command `yum install hadoop-0.20-hive`. The `yum` utility informs you that the Hive package requires `hadoop-0.20` and offers to install it for you. `zypper` and `apt-get` provide similar functionality.

Package Repositories

Package management tools rely on package repositories to install software and resolve any dependency requirements. For information on creating an internal repository, see Using an Internal Package Repository on page 48.

Repository Configuration Files

Information about package repositories is stored in configuration files, the location of which varies according to the package management tool.

- RHEL compatible (`yum`): `/etc/yum.repos.d`
- SLES (`zypper`): `/etc/zypp/zypper.conf`
- Ubuntu (`apt-get`): `/etc/apt/apt.conf` (Additional repositories are specified using `.list` files in the `/etc/apt/sources.list.d/` directory.)

For example, on a typical CentOS system, you might find:

```
ls -l /etc/yum.repos.d/
```

```
total 36
-rw-r--r--. 1 root root 1664 Dec  9  2015 CentOS-Base.repo
-rw-r--r--. 1 root root 1309 Dec  9  2015 CentOS-CR.repo
-rw-r--r--. 1 root root  649 Dec  9  2015 CentOS-Debuginfo.repo
-rw-r--r--. 1 root root  290 Dec  9  2015 CentOS-fasttrack.repo
```

```
-rw-r--r--. 1 root root  630 Dec  9  2015 CentOS-Media.repo
-rw-r--r--. 1 root root 1331 Dec  9  2015 CentOS-Sources.repo
-rw-r--r--. 1 root root 1952 Dec  9  2015 CentOS-Vault.repo
-rw-r--r--. 1 root root  951 Jun 24  2017 epel.repo
-rw-r--r--. 1 root root 1050 Jun 24  2017 epel-testing.repo
```

The `.repo` files contain pointers to one or more repositories. There are similar pointers inside configuration files for `zypper` and `apt-get`. In the following excerpt from `CentOS-Base.repo`, there are two repositories defined: one named `Base` and one named `Updates`. The `mirrorlist` parameter points to a website that has a list of places where this repository can be downloaded.

```
[base]
name=CentOS-$releasever - Base
mirrorlist=http://mirrorlist.centos.org/?release=$releasever&arch=$basearch&repo=os&infra=$infra
#baseurl=http://mirror.centos.org/centos/$releasever/os/$basearch/
gpgcheck=1
gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-CentOS-7

#released updates
[updates]
name=CentOS-$releasever - Updates
mirrorlist=http://mirrorlist.centos.org/?release=$releasever&arch=$basearch&repo=updates&infra=$infra
#baseurl=http://mirror.centos.org/centos/$releasever/updates/$basearch/
gpgcheck=1
gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-CentOS-7
```

Listing Repositories

You can list the enabled repositories by running one of the following commands:

- **RHEL compatible:** `yum repolist`
- **SLES:** `zypper repos`
- **Ubuntu:** `apt-get` does not include a command to display sources, but you can determine sources by reviewing the contents of `/etc/apt/sources.list` and any files contained in `/etc/apt/sources.list.d/`.

The following shows an example of the output of `yum repolist` on a CentOS 7 sytstem:

```
repo id               repo name                                          status
base/7/x86_64         CentOS-7 - Base                                     9,591
epel/x86_64           Extra Packages for Enterprise Linux 7 - x86_64     12,382
extras/7/x86_64       CentOS-7 - Extras                                     392
updates/7/x86_64      CentOS-7 - Updates                                  1,962
repolist: 24,327
```

Using an Internal Parcel Repository

You can create a parcel repository for Cloudera Manager by either hosting an internal repository or by manually copying the repository files to the Cloudera Manager host.

Hosting an Internal Parcel Repository

To host an internal repository, you must install or use an existing web server on an internal host that is reachable by the Cloudera Manager host, and then download the repository files to the web server host.

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 skip to Downloading and Publishing the Parcel Repository on page 46.

An easy web server to install is the Apache HTTP Server.

Install Apache HTTP Server

You may need to respond to some prompts to confirm you want to complete the installation.

OS	Command
RHEL compatible	`sudo yum install httpd`
SLES	`sudo zypper install httpd`
Ubuntu	`sudo apt-get install httpd`

Starting Apache HTTPD

OS	Command
RHEL compatible	`sudo service httpd start`
SLES	`sudo service apache2 start`
Ubuntu	`sudo service apache2 start`

Downloading and Publishing the Parcel Repository

Important: As of February 1, 2021, all downloads of CDH and Cloudera Manager require a username and password and use a modified URL. You must use the modified URL, including the username and password when downloading the repository contents described below. You may need to upgrade Cloudera Manager to a newer version that uses the modified URLs.

This can affect new installations, upgrades, adding new hosts to a cluster, downloading a new parcel, and adding a new cluster.

For more information, see Updating an existing CDH/Cloudera Manager deployment to access downloads with authentication.

1. Download the parcel and `manifest.json` files for your OS distribution from:
   - **CDH 5:** Impala, Kudu, Spark 1, and Search are included in the CDH parcel.
     - CDH - `https://username:password@archive.cloudera.com/p/cdh5/parcels/`
     - Accumulo--`https://username:password@archive.cloudera.com/p/accumulo-c5/parcels/`
     - GPL Extras - `https://username:password@archive.cloudera.com/p/gplextras5/parcels/`
   - **CDS 2 Powered by Apache Spark for CDH 5:**
     - The exact parcel name is dependent on the OS. You can find all the parcels at `https://username:password@archive.cloudera.com/p/spark2/parcels/`.
   - **Cloudera Navigator Key Trustee Server:**
     - Go to the Key Trustee Server download page. Select **Parcels** from the **Package or Parcel** drop-down menu, and click **DOWNLOAD NOW**. This downloads the Key Trustee Server parcels and `manifest.json` files in a `.tar.gz` file. Extract the files with the `tar xvfz filename.tar.gz` command.
   - **Cloudera Navigator Key Trustee KMS and HSM KMS:** (Cloudera Navigator HSM KMS is included in the Key Trustee KMS parcel).
     - Go to the Key Trustee KMS download page. Select **Parcels** from the **Package or Parcel** drop-down menu, and click **DOWNLOAD NOW**. This downloads the Key Trustee KMS parcels and `manifest.json` files in a `.tar.gz` file. Extract the files with the `tar xvfz filename.tar.gz` command.
   - **Sqoop Connectors:**
     - `https://archive.cloudera.com/sqoop-connectors/parcels/`

2. Move the `.parcel` and `manifest.json` files to the web server directory, and modify the file permissions. The hosted `.sha1` files are not used by Cloudera Manager when it obtains the parcels from a web server. Instead, it extracts the SHA1 hashes provided in `manifest.json`. For example:

```
sudo mkdir -p /var/www/html/cloudera-parcels/cdh5/<cdh5_version>/
sudo mv *.parcel* /var/www/html/cloudera-parcels/cdh5/<cdh5_version>/
sudo mv manifest.json /var/www/html/cloudera-parcels/cdh5/<cdh5_version>/
sudo chmod -R ugo+rX /var/www/html/cloudera-parcels/cdh5/<cdh5_version>/
```

   Replace *<cdh5_version>* with your CDH version (for example, `5.14.0`). After moving the files and changing permissions, visit `http://<web_server>/cloudera-parcels/cdh5/<cdh5_version>` to verify that you can access the parcel. If you do not see anything, your web server may have been configured to not show indexes.

Configuring the Cloudera Manager Server to Use the Parcel URL for Hosted Repositories

1. Use one of the following methods to open the parcel settings page:
   - **Navigation bar**
     1. Click in the top navigation bar or click **Hosts** and click the **Parcels** tab.
     2. Click the **Configuration** button.
   - **Menu**
     1. Select **Administration** > **Settings**.
     2. Select **Category** > **Parcels** .
2. In the **Remote Parcel Repository URLs** list, click the addition symbol to open an additional row.
3. Enter the path to the parcel. For example: `http://<web_server>/cdh5/<cdh5_version>`
4. Click **Save Changes** to commit the changes.

Using a Local Parcel Repository

To use a local parcel repository, complete the following steps:

1. Open the Cloudera Manager web UI and navigate to the **Parcels** page.
2. Select **Configuration** and verify that you have a **Local Parcel Repository** path set. By default, the directory is `/opt/cloudera/parcel-repo`.
3. Remove any **Remote Parcel Repository URLs** you are not using, including ones that point to Cloudera archives.
4. Add the parcel you want to use to the local parcel repository directory that you specified.
5. In the command line, navigate to the local parcel repository directory.
6. Create a SHA1 hash for the parcel you added and save it to a file named *<parcel_name>*`.parcel.sha`.

   For example, the following command generates a SHA1 hash for the parcel `CDH-5.14.0-1.cdh5.14.0.p0.24-el7.parcel`:

```
sha1sum CDH-5.14.0-1.cdh5.14.0.p0.24-el7.parcel | awk '{ print $1 }' >
CDH-5.14.0-1.cdh5.14.0.p0.24-el7.parcel.sha
```

7. Change the ownership of the parcel and hash files to `cloudera-scm`:

```
sudo chown cloudera-scm:cloudera-scm *.parcel.*
```

8. In the Cloudera Manager web UI, navigate to the **Parcels** page.
9. Click **Check for New Parcels**.
10. When the new parcel appears, download, distribute, and activate the parcel.

Using an Internal Package Repository

This topic describes how to create an internal package repository and direct hosts in your Cloudera Manager deployment to use that repository. You can create a permanent or temporary repository.

After completing these steps, you can install a specific version of Cloudera Manager or install Cloudera Manager on hosts that are not connected to the Internet.

Creating a Permanent Internal Repository

Install 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 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 48.

Install Apache HTTP Server

You may need to respond to some prompts to confirm you want to complete the installation.

OS	Command
RHEL compatible	`sudo yum install httpd`
SLES	`sudo zypper install httpd`
Ubuntu	`sudo apt-get install httpd`

Starting Apache HTTPD

OS	Command
RHEL compatible	`sudo service httpd start`
SLES	`sudo service apache2 start`
Ubuntu	`sudo service apache2 start`

Downloading the Tarball and Publishing Repository Files

Important: As of February 1, 2021, all downloads of CDH and Cloudera Manager require a username and password and use a modified URL. You must use the modified URL, including the username and password when downloading the repository contents described below. You may need to upgrade Cloudera Manager to a newer version that uses the modified URLs.

This can affect new installations, upgrades, adding new hosts to a cluster, downloading a new parcel, and adding a new cluster.

For more information, see Updating an existing CDH/Cloudera Manager deployment to access downloads with authentication.

1. Download the tarball for your OS distribution from the `repo-as-tarball` archive:
   - **Cloudera Manager 5:** `https://`*`username`*`:`*`password`*`@archive.cloudera.com/p/repo-as-tarball/`
   - **CDH 5:** `https://`*`username`*`:`*`password`*`@archive.cloudera.com/p/cdh5/repo-as-tarball/`

   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, move the files to the web server directory, and modify file permissions. For example, you might use the following commands:

```
tar xvfz cm5.14.0-centos7.tar.gz
sudo mv cm /var/www/html
sudo chmod -R ugo+rX /var/www/html/cm
```

After moving files and changing permissions, visit `http://<web_server>/cm` to verify that you see an index of files. If you do not see anything, your web server may have been configured to not show indexes.

Creating a Temporary Internal 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:
   - **Cloudera Manager 5:** `https://username:password@archive.cloudera.com/p/repo-as-tarball/`
   - **CDH 5:** `https://username:password@archive.cloudera.com/p/cdh5/repo-as-tarball/`

   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:

```
cd /tmp
tar xvfz cm5.14.0-centos7.tar.gz
sudo chmod -R ugo+rX /tmp/cm
```

3. Determine a port that your system is not listening on (for example, port 8900).
4. Start a Python SimpleHTTPServer in the current directory:

```
python -m SimpleHTTPServer 8900
```

```
Serving HTTP on 0.0.0.0 port 8900 ...
```

5. Confirm you can get to this hosted package repository by going to `http://<web_server>:8900/cm` in your browser. You should see links for the hosted files.

Modifying Clients to Use the Internal Repository

Having established the repository, modify the client configuration to use it:

OS	Procedure

RHEL compatible	Create `/etc/yum.repos.d/cloudera-repo.repo` files on cluster hosts with the following content, where *<web_server>* is the hostname of the web server:
	`[cloudera-repo]`
	`name=cloudera-repo`
	`baseurl=http://<web_server>/cm/5`
	`enabled=1`
	`gpgcheck=0`

OS	Procedure
SLES	Use the `zypper` utility to update client system repo information by issuing the following command: `zypper addrepo http://<web_server>/cm <alias>`

Ubuntu	Create `/etc/apt/sources.list.d/cloudera-repo.list` files on all cluster hosts with the following content, where *<web_server>* is the hostname of the web server:
	`deb http://<web_server>/cm <codename> <components>`
	You can find the *<codename>* and *<components>* variables in the `./conf/distributions` file in the repository.
	After creating the `.list` file, run the following command:
	`sudo apt-get update`

Manually Install Cloudera Software Packages

This topic shows how to manually install Cloudera software packages, namely Cloudera Manager and CDH. This is useful for environments where it is not possible to use Cloudera Manager to install the required packages, such as organizations where password-less sudo is not permitted.

Although you can install the software packages manually, Cloudera does not support clusters that are not deployed and managed by Cloudera Manager.

Note: If you choose to install CDH manually using these instructions, you cannot use Cloudera Manager to install additional parcels. This can prevent you from using services that are only available via parcel.

Install the Cloudera Manager Server Packages

1. Install the Cloudera Manager Server packages either on the host where the database is installed, or on a host that has access to the database. This host need not be a host in the cluster that you want to manage with Cloudera Manager. On the Cloudera Manager Server host, type the following commands to install the Cloudera Manager packages.

OS	Command
RHEL, if you have a yum repo configured	`sudo yum install cloudera-manager-daemons cloudera-manager-server`
RHEL,if you're manually transferring RPMs	`sudo yum --nogpgcheck localinstall cloudera-manager-daemons-*.rpm` `sudo yum --nogpgcheck localinstall cloudera-manager-server-*.rpm`
SLES	`sudo zypper install cloudera-manager-daemons cloudera-manager-server`
Ubuntu or Debian	`sudo apt-get install cloudera-manager-daemons cloudera-manager-server`

2. If you are using an Oracle database for Cloudera Manager, increase the Java heap size for Cloudera Manager Server. Edit the `/etc/default/cloudera-scm-server` file on the Cloudera Manager server host. Locate the line that begins with `export CM_JAVA_OPTS` and change the `-Xmx2G` option to `-Xmx4G`.
3. On all hosts, run the following command to install the Cloudera Manager agent:

OS	Command
RHEL, CentOS, Oracle Linux	`sudo yum install cloudera-manager-daemons cloudera-manager-agent`
SLES	`sudo zypper install cloudera-manager-daemons cloudera-manager-agent`
Ubuntu	`sudo apt-get install cloudera-manager-daemons cloudera-manager-agent`

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

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.

   For more information on Agent configuration options, see Agent Configuration File.
5. Start the Agents by running the following command on all hosts:

   RHEL 7, SLES 12, Debian 8, Ubuntu 16.04

```
sudo systemctl start cloudera-scm-agent
```

   If the agent starts without errors, no response displays.

   RHEL 5 or 6, SLES 11, Debian 6 or 7, Ubuntu 12.04, 14.04

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

You should see the following:

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

Installing Cloudera Software Using Tarballs

Before proceeding with this path for a new installation, review Cloudera Installation Guide on page 8. If you are upgrading an existing Cloudera Manager installation, see Upgrading Cloudera Manager and CDH.

In this procedure, you install the Oracle JDK, Cloudera Manager Server, and Cloudera Manager Agent software as tarballs and use Cloudera Manager to automate installation of CDH and managed service software as parcels. For a full discussion of deployment options, see Cloudera Installation Guide on page 8.

> **Important:** Tarball installation of Cloudera Manager is deprecated as of Cloudera Manager 5.9.0 and will be removed in version 6.0.0.

To avoid using system packages, and to use tarballs and parcels instead, follow the instructions in this section.

Before You Begin

Install the Oracle JDK

See Step 2: Install Java Development Kit on page 84.

Install and Configure External Databases

If you are using an external database for services or Cloudera Management Service roles, install and configure it following the instructions in Step 4: Install and Configure Databases on page 87.

Cloudera Manager also requires a database. Prepare the Cloudera Manager Server database as described in Preparing the Cloudera Manager Server Database on page 117.

On CentOS 5 and RHEL 5, Install Python 2.6/2.7 and psycopg2 for Hue

Hue in CDH 5 only works with the operating system's native version of Python *when that version is 2.6 and higher*.

CentOS/RHEL 5 ships with Python 2.4 so you must install Python 2.6 (or Python 2.7) and the Python-PostgreSQL Database Adapter, psycopg2 (not psycopg).

If the Hue server is already installed, you must import the psycopg2 connector into Hue's environment **or** create a symbolic link.

```
## Navigate to Hue within your specific CDH parcel version
cd /opt/cloudera/parcels/`ls -l /opt/cloudera/parcels | grep CDH | tail -1 | awk '{print
 $9}'`/lib/hue/build/env/bin
./python2.6
>>>> import psycopg2
```

or ...

```
cd /opt/cloudera/parcels/`ls -l /opt/cloudera/parcels | grep CDH | tail -1 | awk '{print
 $9}'`/lib/hue/build/env/lib/python2.6/site-packages/
ln -s /usr/lib64/python2.6/site-packages/psycopg2 psycopg2
```

Install the Cloudera Manager Server and Agents

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 on page 72.

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 `cloudera-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 Step 4: Install and Configure Databases on page 87. 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.

Create Parcel Directories

1. On the Cloudera Manager Server host, create a parcel repository directory:

```
$ sudo mkdir -p /opt/cloudera/parcel-repo
```

2. Change the directory ownership to be the username you are using to run Cloudera Manager:

```
$ sudo chown username:groupname /opt/cloudera/parcel-repo
```

where *username* and *groupname* are the user and group names (respectively) you are using to run Cloudera Manager. For example, if you use the default username `cloudera-scm`, you would run the command:

```
$ sudo chown cloudera-scm:cloudera-scm /opt/cloudera/parcel-repo
```

3. On each cluster host, create a parcels directory:

```
$ sudo mkdir -p /opt/cloudera/parcels
```

4. Change the directory ownership to be the username you are using to run Cloudera Manager:

```
$ sudo chown username:groupname /opt/cloudera/parcels
```

where *username* and *groupname* are the user and group names (respectively) you are using to run Cloudera Manager. For example, if you use the default username `cloudera-scm`, you would run the command:

```
$ sudo chown cloudera-scm:cloudera-scm /opt/cloudera/parcels
```

Start the Cloudera Manager Server

> **Important:** When you start the Cloudera Manager Server and Agents, Cloudera Manager assumes you are not already running HDFS and MapReduce. If these services are running:
>
> 1. Shut down HDFS and MapReduce. See Stopping Services (CDH 4) or Stopping CDH Services Using the Command Line (CDH 5) for the commands to stop these services.
> 2. Configure the init scripts to *not* start on boot. Use commands similar to those shown in Configuring init to Start Hadoop System Services (CDH 5), but *disable* the start on boot (for example, `$ sudo chkconfig hadoop-hdfs-namenode off`).
>
> Contact Cloudera Support for help converting your existing Hadoop configurations for use with Cloudera Manager.

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
```

  – 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`.

If the Cloudera Manager Server does not start, see Troubleshooting Installation Problems on page 152.

Start the Cloudera Manager Agents

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`.

Install Package Dependencies

When you install with tarballs and parcels, some services may require additional dependencies that are not provided by Cloudera. On each host, install the required packages:

When you install with tarballs and parcels, some services may require additional dependencies that are not provided by Cloudera. On each host, install the required packages, as determined by running the commands on this page: Package Dependencies.

Start and Log into the Cloudera Manager Admin Console

The Cloudera Manager Server URL takes the following form `http://`*`Server host`*`:`*`port`*, where *Server host* is the fully qualified domain name (FQDN) or IP address of the host where the Cloudera Manager Server is installed, and *port* is the port configured for the Cloudera Manager Server. The default port is 7180.

1. Wait several minutes for the Cloudera Manager Server to start. To observe the startup process, run `tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log` on the Cloudera Manager Server host. If the Cloudera Manager Server does not start, see Troubleshooting Installation Problems on page 152.
2. In a web browser, enter `http://Server host:7180`, where *Server host* is the FQDN or IP address of the host where the Cloudera Manager Server is running.

   The login screen for Cloudera Manager Admin Console displays.
3. Log into Cloudera Manager Admin Console. The default credentials are: **Username:** `admin` **Password:** `admin`. Cloudera Manager does not support changing the `admin` username for the installed account. You can change the password using Cloudera Manager after you run the installation wizard. Although you cannot change the `admin` username, you can add a new user, assign administrative privileges to the new user, and then delete the default `admin` account.
4. After you log in, the **Cloudera Manager End User License Terms and Conditions** page displays. Read the terms and conditions and then select **Yes** to accept them.
5. Click **Continue**.

   The **Welcome to Cloudera Manager** page displays.

Welcome to Cloudera Manager

From the **Welcome to Cloudera Manager** page, you can select the edition of Cloudera Manager to install and, optionally, install a license:

1. Choose which edition to install:
   - Cloudera Express, which does not require a license, but provides a limited set of features.
   - Cloudera Enterprise Cloudera Enterprise Trial, which does not require a license, but expires after 60 days and cannot be renewed.
   - Cloudera Enterprise with one of the following license types:
     - Basic Edition
     - Flex Edition
     - Cloudera Enterprise

   If you choose Cloudera Express or Cloudera Enterprise Cloudera Enterprise Trial, you can upgrade the license at a later time. For more information, see Managing Licenses.
2. If you select Cloudera Enterprise, install a license:
   a. Click the **Select License File** text field.
   b. Browse to the location of your license file, select the file, and then click **Open**.
   c. Click the **Upload** button.
3. Information is displayed indicating what the CDH installation includes. At this point, you can click the **Support** drop-down menu to access online Help or the Support Portal.
4. Click **Continue** to proceed with the installation.

Thank you for choosing Cloudera Manager and CDH

The **Thank you for choosing Cloudera Manager and CDH** page lists the software that is available to be installed. Click **Continue** to proceed with the installation.

Specify hosts for your CDH cluster installation

Choose which hosts will run CDH and managed services

1. Do one of the following depending on whether you are using Cloudera Manager to install software:
   - If you are using Cloudera Manager to install software, search for and choose hosts:

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

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

Note: Unqualified hostnames (short names) must be unique in a Cloudera Manager instance. For example, you cannot have both *host01.example.com* and *host01.standby.example.com* managed by the same Cloudera Manager Server.

You can specify multiple addresses and address ranges by separating them with commas, semicolons, tabs, or blank spaces, or by placing them on separate lines. Use this technique to make more specific searches instead of searching overly wide ranges. The scan results will include all addresses scanned, but only scans that reach hosts running SSH will be selected for inclusion in your cluster by default. If you do not know the IP addresses of all of the hosts, you can enter an address range that spans over unused addresses and then clear the hosts that do not exist (and are not discovered) later in this procedure. However, keep in mind that wider ranges will require more time to scan.

2. Click **Search**. Cloudera Manager identifies the hosts on your cluster to allow you to configure them for services. If there are a large number of hosts on your cluster, wait a few moments to allow them to be discovered and shown in the wizard. If the search is taking too long, you can stop the scan by clicking **Abort Scan**. To find additional hosts, click **New Search**, add the host names or IP addresses and click **Search** again. Cloudera Manager scans hosts by checking for network connectivity. If there are some hosts where you want to install services that are not shown in the list, make sure you have network connectivity between the Cloudera Manager Server host and those hosts. Common causes of loss of connectivity are firewalls and interference from SELinux.
3. Verify that the number of hosts shown matches the number of hosts where you want to install services. Clear host entries that do not exist and clear the hosts where you do not want to install services.

2. Click **Continue**.

The **Select Repository** screen displays.

Select Repository

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

1. Select the repository type to use for the installation. In the **Choose Method** section select one of the following:
   - **Use Parcels (Recommended)**

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

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

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

   a. If you selected **Use Parcels (Recommended)** and you do not see the version you want to install, click the **More Options** button to add the repository URL for your version. Repository URLs for CDH 5 are documented in CDH Download Information. After adding the repository, click **Save Changes** and wait a few seconds for the version to appear. If your Cloudera Manager host uses an HTTP proxy, click the **Proxy Settings** button to configure your proxy.
   b. If you selected **Use Packages**, and the version you want to install is not listed, you can select **Custom Repository** to specify a repository that contains the desired version. Repository URLs for CDH 5 version are documented in CDH Download Information.

3. If you selected **Use Parcels (Recommended)**, specify any **Additional Parcels** you want to install.
4. Select the specific release of the Cloudera Manager Agent you want to install on your hosts. You can choose either the version that matches the Cloudera Manager Server you are currently using (recommended) or specify a different version in a custom repository. If you opted to use custom repositories for installation files, you can provide a GPG key URL that applies for all repositories. Repository and GPG key URLs are documented in CDH Download Information.
5. Click **Continue**.

The **Accept JDK License** screen displays.

Accept JDK License

> **Note:** Cloudera, Inc. acquired Oracle JDK software under the Oracle Binary Code License Agreement. Pursuant to Item D(v)(a) of the SUPPLEMENTAL LICENSE TERMS of the Oracle Binary Code License Agreement, use of JDK software is governed by the terms of the Oracle Binary Code License Agreement. By installing the JDK software, you agree to be bound by these terms. If you do not wish to be bound by these terms, then do not install the Oracle JDK.

To allow Cloudera Manager to automatically install the Oracle JDK on cluster hosts, read the JDK license and check the box labeled **Install Oracle Java SE Development Kit (JDK 7)** if you accept the terms. If you installed your own Oracle JDK version in Step 2: Install Java Development Kit on page 84, leave the box unchecked.

If you allow Cloudera Manager to install the JDK, a second checkbox appears, labeled **Install Java Unlimited Strength Encryption Policy Files**. These policy files are required to enable AES-256 encryption in JDK versions lower than 1.8u161. JDK 1.8u161 and higher enable unlimited strength encryption by default, and do not require policy files.

After reading the license terms and checking the applicable boxes, click **Continue**.

The **Single User Mode** page displays.

Single User Mode

The **Single User Mode** page allows you to enable single user mode. Single user mode is not recommended for most environments. If you need to enable single user mode, check the box labeled **Enable Single User Mode**. Otherwise, leave it unchecked, and click **Continue**. The **Enter Login Credentials** page displays.

Enter Login Credentials

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

   You can modify the default SSH port if necessary.

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

Inspect hosts for correctness

The **Inspect Hosts** page runs the Host Inspector to search for common configuration problems. View the results and address any identified problems. Click the **Run Again** button to update the results after making any changes.

After addressing identified problems, click **Finish**.

This completes the **Cluster Installation** wizard and launches the **Cluster Setup** wizard.

Select Services

The **Select Services** page allows you to select the services you want to install and configure. Make sure that you have the appropriate license key for the services you want to use. You can choose from:

- **Core Hadoop**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, and Hue

- **Core with HBase**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and HBase

- **Core with Impala**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Impala

- **Core with Search**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Solr

- **Core with Spark**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Spark

- **All Services**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, HBase, Impala, Solr, Spark, and Key-Value Store Indexer

- **Custom Services**

  Choose your own services. Services required by chosen services will automatically be included. Flume can be added after your initial cluster has been set up.

To include Cloudera Navigator data management, check the box labeled **Include Cloudera Navigator**.

After selecting the services you want to add, click **Continue**. The **Assign Roles** page displays.

Assign Roles

The **Assign Roles** page suggests role assignments for the hosts in your cluster. You can click on the hostname for a role to select a different host. You can also click the **View By Host** button to see all the roles assigned to a host.

To review the recommended role assignments, see Recommended Cluster Hosts and Role Distribution on page 39.

After assigning all of the roles for your services, click **Continue**. The **Setup Database** page displays.

Setup Database

On the **Setup Database** page, you can enter the database names, usernames, and passwords you created in Step 4: Install and Configure Databases on page 87.

Select the database type and enter the database name, username, and password for each service. Click **Test Connection** to validate the settings. If the connection is successful, a green checkmark and the word **Successful** appears next to each service. If there are any problems, the error is reported next to the service that failed to connect.

After verifying that each connection is successful, click **Continue**. The **Review Changes** page displays.

Review Changes

The **Review Changes** page lists default and suggested settings for several configuration parameters, including data directories.

> **Warning:** Do not place DataNode data directories on NAS devices. When resizing an NAS, block replicas can be deleted, which results in missing blocks.

Review and make any necessary changes, and then click **Continue**. The **First Run Command** page displays.

First Run Command

The **First Run Command** page lists the details of the **First Run** command. You can expand the running commands to view the details of any step, including log files and command output. You can filter the view by selecting **Show All Steps**, **Show Only Failed Steps**, or **Show Running Steps**.

After the **First Run** command completes, click **Continue**.

Congratulations!

The **Congratulations!** page reports the success of the setup wizard. Click **Finish** to complete the wizard. The installation is complete.

Cloudera recommends that you change the default password as soon as possible by clicking the logged-in username at the top right of the home screen and clicking **Change Password**.

(Optional) Change the Cloudera Manager User

After configuring your services, the installation wizard automatically starts the Cloudera Management Service, assuming that it runs using `cloudera-scm`. If you configured this service to run using a user other than `cloudera-scm`, the Cloudera Management Service roles do not start automatically. To change the service configuration to use the user account that you selected:

1. Connect to the Cloudera Manager Admin Console.
2. 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.
3. Click the **Configuration** tab.
4. Use the search box to find the property to change. For example, you might enter "system" to find the **System User** and **System Group** properties.
5. Make any changes required to the System User and System Group to ensure Cloudera Manager uses the proper user accounts.
6. Click **Save Changes**.
7. Start the Cloudera Management Service roles.

Change the Default Administrator Password

As soon as possible, change the default administrator password:

1. Click the logged-in username at the far right of the top navigation bar and select **Change Password**.
2. Enter the current password and a new password twice, and then click **OK**.

Configure Oozie Data Purge Settings

If you added an Oozie service, you can change your Oozie configuration to control when data is purged to improve performance, cut down on database disk usage, or to keep the history for a longer period of time. Limiting the size of the Oozie database can also improve performance during upgrades. See Configuring Oozie Data Purge Settings Using Cloudera Manager.

Test the Installation

You can test the installation following the instructions in Testing the Installation on page 144.

Creating Virtual Images of Cluster Hosts

You can create virtual machine images, such as PXE-boot images, Amazon AMIs, and Azure VM images of cluster hosts with pre-deployed Cloudera software that you can use to quickly spin up virtual machines. These images use parcels to install CDH software. This topic describes the procedures to create images of the Cloudera Manager host and worker host and how to instantiate hosts from those images.

Creating a Pre-Deployed Cloudera Manager Host

To create a Cloudera Manager virtual machine image:

1. Instantiate a virtual machine image (an AMI, if you are using Amazon Web Services) based on a supported operating system and start the virtual machine. See the documentation for your virtualization environment for details.
2. Install Cloudera Manager and configure a database. You can configure either a local or remote database.
3. Wait for the Cloudera Manager Admin console to become active.
4. Log in to the Cloudera Manager Admin console.
5. Download any parcels for CDH or other services managed by Cloudera Manager. Do not distribute or activate the parcels.
6. Log in to the Cloudera Manager server host:
   a. Run the following command to stop the Cloudera Manager service:

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

   b. Run the following command to disable autostarting of the `cloudera-scm-server` service:

   - RHEL6.x, CentOS 6.x and SUSE:

```
chkconfig cloudera-scm-server off
```

   - RHEL 7.x /CentOS 7.x.x:

```
systemctl disable cloudera-scm-server.service
```

   - Ubuntu:

```
update-rc.d -f cloudera-scm-server remove
```

7. Create an image of the Cloudera Manager host.See the documentation for your virtualization environment for details.
8. If you installed the Cloudera Manager database on a remote host, also create an image of the database host.

> **Note:** Ensure that there are no clients using the remote database while creating the image.

Instantiating a Cloudera Manager Image

To create a new Cloudera Manager instance from a virtual machine image:

1. Instantiate the Cloudera Manager image.
2. If the Cloudera Manager database will be hosted on a remote host, also instantiate the database host image.
3. Ensure that the `cloudera-scm-server` service is not running by running the following command on the Cloudera Manager host:

```
service cloudera-scm-server status
```

   If it is running, stop it using the following command:

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

4. On the Cloudera Manager host, create a file named `uuid` in the `/etc/cloudera-scm-server` directory. Add a globally unique identifier to this file using the following command:

```
cat /proc/sys/kernel/random/uuid > /etc/cloudera-scm-server/uuid
```

   The existence of this file informs Cloudera Manager to reinitialize its own unique identifier when it starts.
5. Run the following command to start the Cloudera Manager service:

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

6. Run the following command to enable automatic restart for the `cloudera-scm-server`:
   - RHEL6.x, CentOS 6.x and SUSE:

```
chkconfig cloudera-scm-server on
```

   - RHEL 7.x /CentOS 7.x.x:

```
systemctl enable cloudera-scm-server.service
```

   - Ubuntu:

```
update-rc.d -f cloudera-scm-server defaults
```

Creating a Pre-Deployed Worker Host

1. Instantiate a virtual machine image (an AMI, if you are using Amazon Web Services) based on a supported operating system and start the virtual machine. See the documentation for your virtualization environment for details.
2. Download the parcels required for the worker host from the public parcel repository, or from a repository that you have created and save them to a temporary directory. See Cloudera Manager Version and Download Information.
3. From the same location where you downloaded the parcels, download the *parcel_name*`.parcel.sha1` file for each parcel.
4. Calculate and compare the sha1 of the downloaded parcel to ensure that the parcel was downloaded correctly. For example:

```
sha1sum KAFKA-2.0.2-1.2.0.2.p0.5-el6.parcel | awk '{print $1}' >
KAFKA-2.0.2-1.2.0.2.p0.5-el6.parcel.sha
diff KAFKA-2.0.2-1.2.0.2.p0.5-el6.parcel.sha1 KAFKA-2.0.2-1.2.0.2.p0.5-el6.parcel.sha
```

5. Unpack the parcel:
   a. Create the following directories:

- /opt/cloudera/parcels
- /opt/cloudera/parcel-cache

b. Set the ownership for the two directories you just created so that they are owned by the username that the Cloudera Manager agent runs as.

c. Set the permissions for each directory using the following command:

```
sudo chmod 755 <directory>
```

Note that the contents of these directories will be publicly available and can be safely marked as world-readable.

d. Running as the same user that runs the Cloudera Manager agent, extract the contents of the parcel from the temporary directory using the following command:

```
tar -zxvf <parcel> -C /opt/cloudera/parcels/
```

e. Add a symbolic link from the product name of each parcel to the /opt/cloudera/parcels directory.

For example, to link /opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24 to /opt/cloudera/parcels/**CDH**, use the following command:

```
sudo ln -s /opt/cloudera/parcels/CDH-5.14.0-1.cdh5.14.0.p0.24 /opt/cloudera/parcels/CDH
```

f. Mark the parcels to not be deleted by the Cloudera Manager agent on start up by adding a .dont_delete marker file (this file has no contents) to each subdirectory in the /opt/cloudera/parcels directory. For example:

```
touch /opt/cloudera/parcels/CDH/.dont_delete
```

6. Verify the file exists:

```
ls -l /opt/cloudera/parcels/<parcel>
```

You should see output similar to the following:

```
ls -al /opt/cloudera/parcels/CDH
```

```
total 100
drwxr-xr-x  9 root root  4096 Sep 14 14:53 .
drwxr-xr-x  9 root root  4096 Sep 14 06:34 ..
drwxr-xr-x  2 root root  4096 Sep 12 06:39 bin
-rw-r--r--  1 root root     0 Sep 14 14:53 .dont_delete
drwxr-xr-x 26 root root  4096 Sep 12 05:10 etc
drwxr-xr-x  4 root root  4096 Sep 12 05:04 include
drwxr-xr-x  2 root root 69632 Sep 12 06:44 jars
drwxr-xr-x 37 root root  4096 Sep 12 06:39 lib
drwxr-xr-x  2 root root  4096 Sep 12 06:39 meta
drwxr-xr-x  5 root root  4096 Sep 12 06:39 share
```

7. Install the Cloudera Manager agent. If you have not already done so, Step 1: Configure a Repository for Cloudera Manager on page 82.

8. Create an image of the worker host. See the documentation for your virtualization environment for details.

Instantiating a Worker Host

1. Instantiate the Cloudera worker host image.

2. Edit the following file and set the server_host and server_port properties to reference the Cloudera Manager server host.

3. If necessary perform additional steps to configure TLS/SSL. See Configuring TLS Encryption for Cloudera Manager.

4. Start the agent service:

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

Configuring a Custom Java Home Location

Note: Cloudera strongly recommends installing the JDK at `/usr/java/jdk-version`, which allows Cloudera Manager to auto-detect and use the correct JDK version. If you install the JDK anywhere else, you must follow these instructions to configure Cloudera Manager with your chosen location. The following procedure changes the JDK location for Cloudera Management Services and CDH cluster processes only. It does not affect the JDK used by other non-Cloudera processes, or gateway roles.

Although not recommended, the Oracle Java Development Kit (JDK), which Cloudera services require, may be installed at a custom location if necessary. These steps assume you have already installed the JDK as documented in Step 2: Install Java Development Kit on page 84.

To modify the Cloudera Manager configuration to ensure the JDK can be found:

1. Open the Cloudera Manager Admin Console.
2. In the main navigation bar, click the **Hosts** tab. If you are configuring the JDK location on a specific host only, click the link for that host.
3. Click the **Configuration** tab.
4. Select **Category** > **Advanced**.
5. Set the **Java Home Directory** property to the custom location.
6. Click **Save Changes**.
7. Restart all services.

Creating a CDH Cluster Using a Cloudera Manager Template

You can create a new CDH cluster by exporting a *cluster template* from an existing CDH cluster managed by Cloudera Manager. You can then modify the template and use it to create new clusters with the same configuration on a new set of hosts. Use cluster templates to:

- Duplicate clusters for use in developer, test, and production environments.
- Quickly create a cluster for a specific workload.
- Reproduce a production cluster for testing and debugging.

Follow these general steps to create a template and a new cluster:

1. Export the cluster configuration from the source cluster. The exported configuration is a JSON file that details all of the configurations of the cluster. The JSON file includes an `instantiator` section that contains some values you must provide before creating the new cluster.

   See Exporting the Cluster Configuration on page 67.

2. Set up the hosts for the new cluster by installing Cloudera Manager agents and the JDK on all hosts. For secure clusters, also configure a Kerberos key distribution center (KDC) in Cloudera Manager.

   See Preparing a New Cluster on page 67

3. Create any local repositories required for the cluster.

   See Step 1: Configure a Repository for Cloudera Manager on page 82.

4. Complete the `instantiator` section of the cluster configuration JSON document to create a template.

   See Creating the Template on page 67.

5. Import the cluster template to the new cluster.

   See Importing the Template to a New Cluster on page 71.

Exporting the Cluster Configuration

To create a cluster template, you begin by exporting the configuration from the source cluster. The cluster must be running and managed by Cloudera Manager 5.7 or higher.

To export the configuration:

1. Any Host Templates you have created are used to export the configuration. If you do not want to use those templates in the new cluster, delete them. In Cloudera Manager, go to **Hosts** > **Host Templates** and click **Delete** next to the Host Template you want to delete.
2. Delete any Host Templates created by the Cloudera Manager Installation Wizard. They typically have a name like `Template - 1`).
3. Run the following command to download the JSON configuration file to a convenient location for editing:

```
curl -u admin_username:admin_user_password
"http://Cloudera Manager URL/api/v12/clusters/Cluster name/export" >
path_to_file/file_name.json
```

For example:

```
curl -u adminuser:adminpass
"http://myCluster-1.myDomain.com:7180/api/v12/clusters/Cluster1/export" >
myCluster1-template.json
```

Note: Add the `?exportAutoConfig=true` parameter to the command above to include configurations made by Autoconfiguration. These configurations are included for reference only and are not used when you import the template into a new cluster. For example:

```
curl -u admin_username:admin_user_password
"http://Cloudera Manager URL/api/v12/clusters/Cluster
name/export?exportAutoConfig=true" >
path_to_file/file_name.json
```

Preparing a New Cluster

The new cluster into which you import the cluster template must meet the following requirements:

- Database for Cloudera Manager is installed and configured.
- Cloudera Manager 5.7 or higher is installed and running.
- All required databases for CDH services are installed. See Step 4: Install and Configure Databases on page 87.
- The JDK is installed on all cluster hosts.
- The Cloudera Manager Agent is installed and configured on all cluster hosts.
- If the source cluster uses Kerberos, the new cluster must have KDC properties and privileges configured in Cloudera Manager.
- If the source cluster used *packages* to install CDH and managed services, install those packages manually before importing the template. See Managing Software Installation Using Cloudera Manager.

Creating the Template

To create a template, modify the `instantiator` section of the JSON file you downloaded. Lines that contain the string <changeme> require a value that you must supply. Here is a sample `instantiator` section:

```
  "instantiator" : {
    "clusterName" : "<changeme>",
    "hosts" : [ {
      "hostName" : "<changeme>",
      "hostTemplateRefName" : "<changeme>",
      "roleRefNames" : [ "HDFS-1-NAMENODE-0be88b55f5dedbf7bc74d61a86c0253e" ]
    }, {
```

```
    "hostName" : "<changeme>",
    "hostTemplateRefName" : "<changeme>"
  }, {
    "hostNameRange" : "<HOST[0001-0002]>",
    "hostTemplateRefName" : "<changeme>"
  } ],
  "variables" : [ {
    "name" : "HDFS-1-NAMENODE-BASE-dfs_name_dir_list",
    "value" : "/dfs/nn"
  }, {
    "name" : "HDFS-1-SECONDARYNAMENODE-BASE-fs_checkpoint_dir_list",
    "value" : "/dfs/snn"
  }, {
    "name" : "HIVE-1-hive_metastore_database_host",
    "value" : "myCluster-1.myDomain.com"
  }, {
    "name" : "HIVE-1-hive_metastore_database_name",
    "value" : "hive1"
  }, {
    "name" : "HIVE-1-hive_metastore_database_password",
    "value" : "<changeme>"
  }, {
    "name" : "HIVE-1-hive_metastore_database_port",
    "value" : "3306"
  }, {
    "name" : "HIVE-1-hive_metastore_database_type",
    "value" : "mysql"
  }, {
    "name" : "HIVE-1-hive_metastore_database_user",
    "value" : "hive1"
  }, {
    "name" : "HUE-1-database_host",
    "value" : "myCluster-1.myDomain.com"
  }, {
    "name" : "HUE-1-database_name",
    "value" : "hueserver0be88b55f5dedbf7bc74d61a86c0253e"
  }, {
    "name" : "HUE-1-database_password",
    "value" : "<changeme>"
  }, {
    "name" : "HUE-1-database_port",
    "value" : "3306"
  }, {
    "name" : "HUE-1-database_type",
    "value" : "mysql"
  }, {
    "name" : "HUE-1-database_user",
    "value" : "hueserver0be88b5"
  }, {
    "name" : "IMPALA-1-IMPALAD-BASE-scratch_dirs",
    "value" : "/impala/impalad"
  }, {
    "name" : "KUDU-1-KUDU_MASTER-BASE-fs_data_dirs",
    "value" : "/var/lib/kudu/master"
  }, {
    "name" : "KUDU-1-KUDU_MASTER-BASE-fs_wal_dir",
    "value" : "/var/lib/kudu/master"
  }, {
    "name" : "KUDU-1-KUDU_TSERVER-BASE-fs_data_dirs",
    "value" : "/var/lib/kudu/tserver"
  }, {
    "name" : "KUDU-1-KUDU_TSERVER-BASE-fs_wal_dir",
    "value" : "/var/lib/kudu/tserver"
  }, {
    "name" : "MAPREDUCE-1-JOBTRACKER-BASE-jobtracker_mapred_local_dir_list",
    "value" : "/mapred/jt"
  }, {
    "name" : "MAPREDUCE-1-TASKTRACKER-BASE-tasktracker_mapred_local_dir_list",
    "value" : "/mapred/local"
  }, {
    "name" : "OOZIE-1-OOZIE_SERVER-BASE-oozie_database_host",
    "value" : "myCluster-1.myDomain.com:3306"
  }, {
```

```
      "name" : "OOZIE-1-OOZIE_SERVER-BASE-oozie_database_name",
      "value" : "oozieserver0be88b55f5dedbf7bc74d61a86c0253e"
    }, {
      "name" : "OOZIE-1-OOZIE_SERVER-BASE-oozie_database_password",
      "value" : "<changeme>"
    }, {
      "name" : "OOZIE-1-OOZIE_SERVER-BASE-oozie_database_type",
      "value" : "mysql"
    }, {
      "name" : "OOZIE-1-OOZIE_SERVER-BASE-oozie_database_user",
      "value" : "oozieserver0be88"
    }, {
      "name" : "YARN-1-NODEMANAGER-BASE-yarn_nodemanager_local_dirs",
      "value" : "/yarn/nm"
    }, {
      "name" : "YARN-1-NODEMANAGER-BASE-yarn_nodemanager_log_dirs",
      "value" : "/yarn/container-logs"
    } ]
  }
```

To modify the template:

1. Update the `hosts` section.

   If you have host templates defined in the source cluster, they appear in the `hostTemplates` section of the JSON template. For hosts that do not use host templates, the export process creates host templates based on role assignments to facilitate creating the new cluster. In either case, you must match the items in the `hostTemplates` section with the `hosts` sections in the `instantiator` section.

   Here is a sample of the `hostTemplates` section from the same JSON file as the `instantiator` section, above:

```
 "hostTemplates" : [ {
    "refName" : "HostTemplate-0-from-myCluster-1.myDomain.com",
    "cardinality" : 1,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-GATEWAY-BASE",
"HBASE-1-HBASETHRIFTSERVER-BASE", "HBASE-1-MASTER-BASE", "HDFS-1-BALANCER-BASE",
"HDFS-1-GATEWAY-BASE", "HDFS-1-NAMENODE-BASE", "HDFS-1-NFSGATEWAY-BASE",
"HDFS-1-SECONDARYNAMENODE-BASE", "HIVE-1-GATEWAY-BASE", "HIVE-1-HIVEMETASTORE-BASE",
"HIVE-1-HIVESERVER2-BASE", "HUE-1-HUE_LOAD_BALANCER-BASE", "HUE-1-HUE_SERVER-BASE",
"IMPALA-1-CATALOGSERVER-BASE", "IMPALA-1-STATESTORE-BASE", "KAFKA-1-KAFKA_BROKER-BASE",
 "KS_INDEXER-1-HBASE_INDEXER-BASE", "KUDU-1-KUDU_MASTER-BASE", "MAPREDUCE-1-GATEWAY-BASE",
 "MAPREDUCE-1-JOBTRACKER-BASE", "OOZIE-1-OOZIE_SERVER-BASE", "SOLR-1-SOLR_SERVER-BASE",
 "SPARK_ON_YARN-1-GATEWAY-BASE", "SPARK_ON_YARN-1-SPARK_YARN_HISTORY_SERVER-BASE",
"SQOOP-1-SQOOP_SERVER-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-GATEWAY-BASE",
"YARN-1-JOBHISTORY-BASE", "YARN-1-RESOURCEMANAGER-BASE", "ZOOKEEPER-1-SERVER-BASE" ]
  }, {
    "refName" : "HostTemplate-1-from-myCluster-4.myDomain.com",
    "cardinality" : 1,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-REGIONSERVER-BASE",
"HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IMPALAD-BASE",
"KUDU-1-KUDU_TSERVER-BASE", "MAPREDUCE-1-TASKTRACKER-BASE",
"SPARK_ON_YARN-1-GATEWAY-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-NODEMANAGER-BASE"
 ]
  }, {
    "refName" : "HostTemplate-2-from-myCluster-[2-3].myDomain.com",
    "cardinality" : 2,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-REGIONSERVER-BASE",
"HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IMPALAD-BASE",
"KAFKA-1-KAFKA_BROKER-BASE", "KUDU-1-KUDU_TSERVER-BASE", "MAPREDUCE-1-TASKTRACKER-BASE",
 "SPARK_ON_YARN-1-GATEWAY-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-NODEMANAGER-BASE"
 ]
  } ]
```

   The value of `cardinality` indicates how many hosts are assigned to the host template in the source cluster.

   The value of `roleConfigGroupsRefNames` indicates which role groups are assigned to the host(s).

   Do the following for each host template in the `hostTemplates` section:

   1. Locate the entry in the `hosts` section of the `instantiator` where you want the roles to be installed.
   2. Copy the value of the `refName` to the value for `hostTemplateRefName`.

3. Enter the hostname in the new cluster as the value for `hostName`. Some host sections might instead use `hostNameRange` for clusters with multiple hosts that have the same set of roles. Indicate a range of hosts by using one of the following:
   - Brackets; for example, `myhost[1-4].foo.com`
   - A comma-delimited string of hostnames; for example, `host-1.domain, host-2.domain, host-3.domain`

Here is an example of the `hostTemplates` and the `hosts` section of the `instantiator` completed correctly:

```
"hostTemplates" : [ {
    "refName" : "HostTemplate-0-from-myCluster-1.myDomain.com",
    "cardinality" : 1,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-GATEWAY-BASE",
"HBASE-1-HBASETHRIFTSERVER-BASE", "HBASE-1-MASTER-BASE", "HDFS-1-BALANCER-BASE",
"HDFS-1-GATEWAY-BASE", "HDFS-1-NAMENODE-BASE", "HDFS-1-NFSGATEWAY-BASE",
"HDFS-1-SECONDARYNAMENODE-BASE", "HIVE-1-GATEWAY-BASE", "HIVE-1-HIVEMETASTORE-BASE",
"HIVE-1-HIVESERVER2-BASE", "HUE-1-HUE_LOAD_BALANCER-BASE", "HUE-1-HUE_SERVER-BASE",
"IMPALA-1-CATALOGSERVER-BASE", "IMPALA-1-STATESTORE-BASE", "KAFKA-1-KAFKA_BROKER-BASE",
 "KS_INDEXER-1-HBASE_INDEXER-BASE", "KUDU-1-KUDU_MASTER-BASE", "MAPREDUCE-1-GATEWAY-BASE",
 "MAPREDUCE-1-JOBTRACKER-BASE", "OOZIE-1-OOZIE_SERVER-BASE", "SOLR-1-SOLR_SERVER-BASE",
 "SPARK_ON_YARN-1-GATEWAY-BASE", "SPARK_ON_YARN-1-SPARK_YARN_HISTORY_SERVER-BASE",
"SQOOP-1-SQOOP_SERVER-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-GATEWAY-BASE",
"YARN-1-JOBHISTORY-BASE", "YARN-1-RESOURCEMANAGER-BASE", "ZOOKEEPER-1-SERVER-BASE" ]
  }, {
    "refName" : "HostTemplate-1-from-myCluster-4.myDomain.com",
    "cardinality" : 1,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-REGIONSERVER-BASE",
"HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IMPALAD-BASE",
"KUDU-1-KUDU_TSERVER-BASE", "MAPREDUCE-1-TASKTRACKER-BASE",
"SPARK_ON_YARN-1-GATEWAY-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-NODEMANAGER-BASE"
 ]
  }, {
    "refName" : "HostTemplate-2-from-myCluster-[2-3].myDomain.com",
    "cardinality" : 2,
    "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-REGIONSERVER-BASE",
"HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IMPALAD-BASE",
"KAFKA-1-KAFKA_BROKER-BASE", "KUDU-1-KUDU_TSERVER-BASE", "MAPREDUCE-1-TASKTRACKER-BASE",
 "SPARK_ON_YARN-1-GATEWAY-BASE", "SQOOP_CLIENT-1-GATEWAY-BASE", "YARN-1-NODEMANAGER-BASE"
 ]
  } ],
  "instantiator" : {
    "clusterName" : "myCluster_new",
    "hosts" : [ {
      "hostName" : "myNewCluster-1.myDomain.com",
      "hostTemplateRefName" : "HostTemplate-0-from-myCluster-1.myDomain.com",
      "roleRefNames" : [ "HDFS-1-NAMENODE-c975a0b51fd36e914896cd5e0adb1b5b" ]
    }, {
      "hostName" : "myNewCluster-5.myDomain.com",
      "hostTemplateRefName" : "HostTemplate-1-from-myCluster-4.myDomain.com"
    }, {
      "hostNameRange" : "myNewCluster-[3-4].myDomain.com",
      "hostTemplateRefName" : "HostTemplate-2-from-myCluster-[2-3].myDomain.com"
    } ],
```

2. For host sections that have a `roleRefNames` line, determine the role type and assign the appropriate host for the role. If there are multiple instances of a role, you must select the correct hosts. To determine the role type, search the template file for the value of `roleRefNames`.

   For example: For a role ref named `HDFS-1-NAMENODE-0be88b55f5dedbf7bc74d61a86c0253e`, if you search for that string, you find a section similar to the following:

```
"roles": [
{
"refName": "HDFS-1-NAMENODE-0be88b55f5dedbf7bc74d61a86c0253e",
"roleType": "NAMENODE"
}
]
```

In this case, the role type is NAMENODE.

3. Modify the `variables` section. This section contains various properties from the source cluster. You can change any of these values to be different in the new cluster, or you can leave the values as copied from the source. For any values shown as <changeme>, you must provide the correct value.

> **Note:** Many of these variables contain information about databases used by the Hive Metastore and other CDH components. Change the values of these variables to match the databases configured for the new cluster.

4. Enter the internal name of the new cluster on the line with `"clusterName" : "<changeme>"`. For example:

```
"clusterName" : "QE_test_cluster"
```

5. (Optional) Change the display name for the cluster. Edit the line that begins with `"displayName"` (near the top of the JSON file); for example:

```
"displayName" : "myNewCluster",
```

Importing the Template to a New Cluster

To import the cluster template:

1. Log in to the Cloudera Manager server as root.
2. Run the following command to import the template. If you have remote repository URLS configured in the source cluster, append the command with `?addRepositories=true`.

```
curl -X POST -H "Content-Type: application/json" -d
      @path_to_template/template_filename.json
http://admin_user:admin_password@cloudera_manager_url:cloudera_manager_port/api/v12/cm/importClusterTemplate
```

You should see a response similar to the following:

```
{
  "id" : 17,
  "name" : "ClusterTemplateImport",
  "startTime" : "2016-03-09T23:44:38.491Z",
  "active" : true,
  "children" : {
    "items" : [ ]
  }
```

Examples:

```
curl -X POST -H "Content-Type: application/json" -d @myTemplate.json
http://admin:admin@myNewCluster-1.mydomain.com:7182/api/v12/cm/importClusterTemplate
```

```
curl -X POST -H "Content-Type: application/json" -d @myTemplate.json
http://admin:admin@myNewCluster-1.mydomain.com:7182/api/v12/cm/importClusterTemplate?addRepositories=true
```

If there is no response, or you receive an error message, the JSON file may be malformed, or the template may have invalid hostnames or invalid references. Inspect the JSON file, correct any errors, and then re-run the command.

3. Open Cloudera Manager for the new cluster in a web browser and click the Cloudera Manager logo to go to the home page.
4. Click the **All Recent Commands** tab.

If the import is proceeding, you should see a link labeled **Import Cluster Template**. Click the link to view the progress of the import.

If any of the commands fail, correct the problem and click **Retry**. You may need to edit some properties in Cloudera Manager.

After you import the template, Cloudera Manager applies the Autoconfiguration rules that set properties such as memory and CPU allocations for various roles. If the new cluster has different hardware or operational requirements, you may need to modify these values.

Sample Python Code

You can perform the steps to export and import a cluster template programmatically using a client written in Python or other languages. (You can also use the `curl` commands provided above.)

Python export example:

```
resource = ApiResource("myCluster-1.myDomain.com", 7180, "admin", "admin", version=12)
cluster = resource.get_cluster("Cluster1");
template = cluster.export(False)
pprint(template)
```

Python import example:

```
resource = ApiResource("localhost", 8180, "admin", "admin", version=12)
 with open('~/cluster-template.json') as data_file:
   data = json.load(data_file)
template = ApiClusterTemplate(resource).from_json_dict(data, resource)
cms = ClouderaManager(resource)
cms.import_cluster_template(template)
```

Configuring Single User Mode

In a conventional Cloudera Manager deployment, the Cloudera Manager Agent, which manages Hadoop processes on each host, runs as the root user. However, some environments restrict access to the root account.

Cloudera Manager 5.3 and higher provides ***single user mode***, which satisfies the requirements of such environments. In single user mode, the Cloudera Manager Agent and *all the processes run by services managed by Cloudera Manager* are started as a single configured user and group. Single user mode prioritizes isolation between Hadoop and the rest of the system over isolation between Hadoop processes running on the system.

Within a Cloudera Manager deployment, single user mode is global and applies to all clusters managed by that instance of Cloudera Manager.

By default, the single user is `cloudera-scm` and the configuration steps described in the following sections assume that user. However, other users are supported. If you choose another user, replace `cloudera-scm` in the following steps with the selected user, and perform the additional steps in Using a Non-default Single User on page 73.

The following sections describe limitations of single user mode and the required configuration steps for the supported installation scenarios at specific points during the installation process.

Limitations

- Switching between conventional and single user mode is not supported.

  Important: In **Administration** > **Settings** there is a **Single user mode** checkbox. Do not enable or disable the checkbox after installation.

- Single user mode is supported for clusters running CDH 5.2 and higher.
- NFS Gateway is not supported in single user mode.
- Cloudera Navigator data encryption components are not supported in single user mode.
- Kudu is not supported in single user mode.

Using a Non-default Single User

When configuring single user mode for a user other than the default (`cloudera-scm`), perform the following configuration steps:

- Make the following directories writable by the single user:
  - `/var/log/cloudera-scm-agent/`
  - `/var/lib/cloudera-scm-agent/`
- Cloudera Manager stores parcels under `/opt/cloudera`, which by default is owned by `cloudera-scm`. Do one of the following:
  - Change `/opt/cloudera` to be writable by the single user.
  - Change the parcel directory location to be writable by the single user:
    1. Go to **Administration** > **Settings** > **Parcels**.
    2. Set the **Local Parcel Repository Path** property.
    3. Click **Save Changes**.
- For a single user *username*, create the process limits configuration file at `/etc/security/limits.d/username.conf` with the following settings:

```
username soft nofile 32768
username soft nproc 65536
username hard nofile 1048576
username hard nproc unlimited
username hard memlock unlimited
username soft memlock unlimited
```

Configuration Steps Before Starting Cloudera Manager Agents in Installation Paths B and C

- If you manually install Agent packages, before starting the Agents, configure them to run as `cloudera-scm` by editing the file `/etc/default/cloudera-scm-agent` and uncommenting the line:

```
USER="cloudera-scm"
```

- Configure the parcels directory. Do one of the following:
  - On each host, in the Agent configuration file `/etc/cloudera-scm-agent/config.ini`, set the `parcel_dir` property:

```
# Parcel directory.  Unpacked parcels will be stored in this directory.
# Downloaded parcels will be stored in <parcel_dir>/../parcel-cache
# parcel_dir=/opt/cloudera/parcels
```

  - 1. Click **Hosts** in the top navigation bar.
    2. Click the **Configuration** tab.
    3. Select **Category** > **Parcels**.
    4. Configure the value of the **Parcel Directory** property. The setting of the `parcel_dir` property in the Cloudera Manager Agent configuration file overrides this setting.
    5. Click **Save Changes** to commit the changes.
    6. Restart the Cloudera Manager Agent on all hosts.

Configuration Steps Before Running the Installation Wizard

Before configuring a cluster to run in single user mode, the following steps must be performed on *all hosts in the cluster*:

- Give the single user passwordless sudo access. You must create the user if it doesn't exist. One common way of achieving this is to add the user to the configured sudoers group by running the command:

```
usermod -a -G sudo cloudera-scm
```

  or adding a new sudo configuration for the `cloudera-scm` group by running the command `visudo` and then adding the following line:

```
%cloudera-scm ALL=(ALL) NOPASSWD: ALL
```

- Sudo must be configured so that `/usr/sbin` is in the path when running sudo. One way to achieve this is by adding the following configuration to sudoers:
  1. Edit the `/etc/sudoers` file using the `visudo` command
  2. Add this line to the configuration file:

```
Defaults secure_path = /sbin:/bin:/usr/sbin:/usr/bin
```

- Set up per user limits for su prior to setting up the Agent.
  1. Edit`/etc/pam.d/su`.
  2. Uncomment:

```
session required pam_limits.so
```

- Roles that run on Tomcat require some directories to exist in non-configurable paths. The following directories must be created and be writable by `cloudera-scm`:
  - **HDFS** (HttpFS role) - `/var/lib/hadoop-httpfs`
  - **Oozie Server** - `/var/lib/oozie`
  - **Sqoop 2 Server** - `/var/lib/sqoop2`
  - **Solr Server** - `/var/lib/solr`
- Cloudera recommends that you create a prefix directory (for example, `/cm`) owned by `cloudera-scm` under which all other service directories will be placed. In single user mode, the Cloudera Manager Agent creates directories under the prefix directory with the correct ownership. If hosts have additional volumes on them that will be used for data directories Cloudera recommends creating a directory on each volume (for example, `/data0/cm` and `/data1/cm`) that is writable by `cloudera-scm`.

Configuration Steps Before Starting the Installation Wizard in Installation Paths B and C

Perform the following steps for the indicated scenarios:

- **Path C** - Do one of the following:
  - Create and change the ownership of `/var/lib/cloudera-scm-server` to the single user.
  - Set the Cloudera Manager Server local storage directory to one owned by the single user:
    1. Go to **Administration** > **Settings** > **Advanced**.
    2. Set the **Cloudera Manager Server Local Data Storage Directory** property to a directory owned by the single user.
    3. Click **Save Changes** to commit the changes.
- **Path B and C when using already managed hosts** - Configure single user mode:
  1. Go to **Administration** > **Settings** > **Advanced**.
  2. Check the **Single User Mode** checkbox.
  3. Click **Save Changes** to commit the changes.

Configuration Steps While Running the Installation Wizard

When configuring the first cluster in Cloudera Manager using the Installation wizard you'll have the option to set up the cluster in single user mode. This configures the Agents to run as `cloudera-scm`.

During the review configuration step you confirm that all the configured paths are writable by `cloudera-scm`. The directories themselves don't have to exist as long as the parent directory is writable by `cloudera-scm`.

Following the standard review configuration page, an additional paths configuration page shows all the configurable paths for the services that will be created in the cluster. These must also be modified to be locations writable by `cloudera-scm`. In most cases, the paths that need to be modified from their default locations fall under two categories:

- Paths under `/var` - These are `log`, `run`, and `data` directories for the different services.
- Per volume data directories - These are data directory configurations that list a directory per volume. Such configurations are used by HDFS, MapReduce, YARN and Impala.

Configuration for Secure Clusters

You must perform some additional configuration when setting up secure HDFS in single user mode:

- When configuring Kerberos, also refer to Kerberos Authentication for Single User Mode and Non-Default Users.
- Configure HDFS with TLS/SSL encryption.
- Do not configure the DataNode Transceiver port and HTTP Web UI port to use privileged ports.
- Configure DataNode data transfer protection.

Controlling Access to sudo Commands

To comply with security requirements, you may need to control access to the sudo commands run by Cloudera Manager Agents. You can control access by creating a "whitelist" of sudo commands that the Cloudera Manager Agent runs, or you can override the `sudo` command so that a custom script that denies some actions is called instead.

Cloudera Manager Agents use `sudo` for the following regular operations:

- Running the `update-alternatives` command during upgrades and when updating parcels.
- Adding new roles or services that require `sudo` access to the `chown` and `chmod` commands.
- Running the `mount` and `unmount` commands when performing resource management activities that use Linux Control Groups (cgroups) and mounting a `tmpfs` mount point for temporary directories for cm_processes.
- Collecting diagnostics, which requires reading files from the `/proc` and `/etc` directories and distribution-specific networking configuration files.

Whitelisting sudo Commands

The following commands may need to be whitelisted for the `cloudera-scm-agent` user. This can be either the default user, `cloudera-scm`, or a single user you specify. Use Linux tools to manage access to these commands as required for your deployment. (See the `man` page for `sudoers`.)

Regular Operation Commands

- `cat`
- `chmod`
- `chown`
- `cp`
- `find`
- `mkdir`
- `mount`
- `rm`
- `umount`
- `update-alternatives`

Support Bundle Collection Commands

- `bash`

- `cat`
- `date`
- `df`
- `dmesg`
- `grep`
- `hostname`
- `ls`
- `netstat`
- `ps`
- `rpm`
- `uname`
- `chkconfig`
- `ethtool`
- `ifconfig`
- `iptables`
- `lsmod`
- `lspci`
- `lvdisplay`
- `service`
- `sysctl`
- `curl`
- `dig`
- `host`
- `lsb_release`
- `lscpu`
- `nslookup`
- `ntpstat`
- `python`
- `sar`
- `top`
- `uptime`
- `vmstat`
- `dmidecode`
- `lsof`
- `ntpq`

Overriding the sudo Command

You can override the `sudo` command so that a custom script is called instead. This script can deny some actions.

To configure the location of this script:

1. Edit the `/etc/cloudera-scm-agent/config.ini` file *on all cluster hosts* and add the following line:

```
sudo_command=path_to_script
```

2. Restart the Cloudera Manager Agent on all cluster hosts:

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

To help determine which commands to override, see the following samples of typical commands run by Cloudera Manager Agent.

Commands run by the Cloudera Manager Agent while it brings up roles for new services on a single host:

```
/bin/cat /proc/cgroups
/bin/chmod -R ugo+r /etc/accumulo/*
/bin/chmod -R ugo+r /etc/hadoop/*
/bin/chmod -R ugo+r /etc/hbase/*
/bin/chmod -R ugo+r /etc/hive/*
/bin/chmod -R ugo+r /etc/solr/*
/bin/chmod -R ugo+r /etc/spark/*
/bin/chmod -R ugo+r /etc/sqoop/*
/bin/chmod 0644 /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/topology.map
/bin/chmod 0755 /cldr/app/coolapp/opt/parcels/CDH*/lib/hue/desktop
/bin/chmod 0755 /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/topology.py
/bin/chmod 4754
/cldr/app/coolapp/opt/parcels/CDH*/lib/hadoop-0.20-mapreduce/sbin/Linux/task-controller
/bin/chmod 6050 /cldr/app/coolapp/opt/parcels/CDH*/lib/hadoop-yarn/bin/container-executor
/bin/chown -R cloudera-scm:cloudera-scm /tmp/*
/bin/chown coolapp:coolapp /etc/hadoop/*/topology.map
/bin/chown coolapp:coolapp /etc/hadoop/*/topology.py
/bin/chown coolapp:coolapp /cldr/app/coolapp/var/run/coolapp-agent/process/*/topology.map
/bin/chown coolapp:coolapp /cldr/app/coolapp/var/run/coolapp-agent/process/*/topology.py
/bin/chown root /etc/accumulo/*
/bin/chown root /etc/hadoop/*
/bin/chown root /etc/hbase/*
/bin/chown root /etc/hive/*
/bin/chown root /etc/solr/*
/bin/chown root /etc/spark/*
/bin/chown root /etc/sqoop/*
/bin/chown root
/cldr/app/coolapp/opt/parcels/CDH*/lib/hadoop-0.20-mapreduce/sbin/Linux/task-controller
/bin/chown root /cldr/app/coolapp/opt/parcels/CDH*/lib/hadoop-yarn/bin/container-executor
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/accumulo-conf
/etc/accumulo/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hadoop-conf
/etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/yarn-conf /etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hadoop-conf
/etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/yarn-conf /etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hadoop-conf
/etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hadoop-conf
/etc/hadoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hbase-conf /etc/hbase/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hbase-conf /etc/hbase/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hive-conf /etc/hive/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/hive-conf /etc/hive/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/solr-conf /etc/solr/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/solr-conf /etc/solr/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/spark-conf /etc/spark/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/spark-conf /etc/spark/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/sqoop-conf /etc/sqoop/*
/bin/cp -a /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/sqoop-conf /etc/sqoop/*
/bin/cp -p /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/topology.map
/etc/hadoop/*/topology.map
/bin/cp -p /cldr/app/coolapp/var/run/cloudera-scm-agent/process/*/topology.py
/etc/hadoop/*/topology.py
/bin/mkdir -p /etc/accumulo
/bin/mkdir -p /etc/flume-ng
/bin/mkdir -p /etc/hadoop
/bin/mkdir -p /etc/hadoop-httpfs
/bin/mkdir -p /etc/hadoop-kms
/bin/mkdir -p /etc/hbase
/bin/mkdir -p /etc/hbase-solr
/bin/mkdir -p /etc/hive
/bin/mkdir -p /etc/hive-hcatalog
/bin/mkdir -p /etc/hive-webhcat
/bin/mkdir -p /etc/hue
/bin/mkdir -p /etc/impala
/bin/mkdir -p /etc/llama
/bin/mkdir -p /etc/mahout
/bin/mkdir -p /etc/oozie
/bin/mkdir -p /etc/pig
/bin/mkdir -p /etc/sentry
```

```
/bin/mkdir -p /etc/solr
/bin/mkdir -p /etc/spark
/bin/mkdir -p /etc/sqoop
/bin/mkdir -p /etc/sqoop2
/bin/mkdir -p /etc/zookeeper
/bin/mount -t cgroup -o blkio cm_cgroups /tmp/*
/bin/mount -t cgroup -o cpu cm_cgroups /tmp/*
/bin/mount -t cgroup -o cpuacct cm_cgroups /tmp/*
/bin/mount -t cgroup -o memory cm_cgroups /tmp/*
/bin/mount -t tmpfs cm_processes /cldr/app/coolapp/var/run/cloudera-scm-agent/process
-o mode
/bin/rm
/bin/rm -rf /etc/accumulo/*
/bin/rm -rf /etc/hadoop/*
/bin/rm -rf /etc/hbase/*
/bin/rm -rf /etc/hive/*
/bin/rm -rf /etc/solr/*
/bin/rm -rf /etc/spark/*
/bin/rm -rf /etc/sqoop/*
/bin/umount /tmp/*
/usr/sbin/update-alternatives --admindir /var/lib/alternatives --altdir /etc/alternatives
 --display ip6tables.x86_64
/usr/sbin/update-alternatives --admindir /var/lib/alternatives --altdir /etc/alternatives
 --display iptables.x86_64
/usr/sbin/update-alternatives --admindir /var/lib/alternatives --altdir /etc/alternatives
 --display mta
/usr/sbin/update-alternatives --admindir /var/lib/alternatives --altdir /etc/alternatives
 --display print
/usr/sbin/update-alternatives --auto accumulo-conf
/usr/sbin/update-alternatives --auto hadoop-conf
/usr/sbin/update-alternatives --auto hbase-conf
/usr/sbin/update-alternatives --auto hive-conf
/usr/sbin/update-alternatives --auto solr-conf
/usr/sbin/update-alternatives --auto spark-conf
/usr/sbin/update-alternatives --auto sqoop-conf
/usr/sbin/update-alternatives --install /etc/accumulo/conf accumulo-conf /etc/accumulo/*
 51
/usr/sbin/update-alternatives --install /etc/flume-ng/conf flume-ng-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/flume-ng/conf.empty 10
/usr/sbin/update-alternatives --install /etc/hadoop-httpfs/conf hadoop-httpfs-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hadoop-httpfs/conf.empty 10
/usr/sbin/update-alternatives --install /etc/hadoop-kms/conf hadoop-kms-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hadoop-kms/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hadoop/conf hadoop-conf /etc/hadoop/* 90
/usr/sbin/update-alternatives --install /etc/hadoop/conf hadoop-conf /etc/hadoop/* 91
/usr/sbin/update-alternatives --install /etc/hadoop/conf hadoop-conf /etc/hadoop/* 92
/usr/sbin/update-alternatives --install /etc/hadoop/conf hadoop-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hadoop/conf.empty 10
/usr/sbin/update-alternatives --install /etc/hbase-solr/conf hbase-solr-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hbase-solr/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hbase/conf hbase-conf /etc/hbase/* 90
/usr/sbin/update-alternatives --install /etc/hbase/conf hbase-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hbase/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hive-hcatalog/conf hive-hcatalog-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hive-hcatalog/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hive-webhcat/conf hive-webhcat-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hive-webhcat/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hive/conf hive-conf /etc/hive/* 90
/usr/sbin/update-alternatives --install /etc/hive/conf hive-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hive/conf.dist 10
/usr/sbin/update-alternatives --install /etc/hue/conf hue-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/hue/conf.empty 10
/usr/sbin/update-alternatives --install /etc/impala/conf impala-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/impala/conf.dist 10
/usr/sbin/update-alternatives --install /etc/llama/conf llama-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/llama/conf.dist 10
/usr/sbin/update-alternatives --install /etc/mahout/conf mahout-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/mahout/conf.dist 10
/usr/sbin/update-alternatives --install /etc/oozie/conf oozie-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/oozie/conf.dist 10
/usr/sbin/update-alternatives --install /etc/pig/conf pig-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/pig/conf.dist 10
/usr/sbin/update-alternatives --install /etc/sentry/conf sentry-conf
```

```
/cldr/app/coolapp/opt/parcels/CDH*/etc/sentry/conf.dist 10
/usr/sbin/update-alternatives --install /etc/solr/conf solr-conf /etc/solr/* 90
/usr/sbin/update-alternatives --install /etc/solr/conf solr-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/solr/conf.dist 10
/usr/sbin/update-alternatives --install /etc/spark/conf spark-conf /etc/spark/* 51
/usr/sbin/update-alternatives --install /etc/spark/conf spark-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/spark/conf.dist 10
/usr/sbin/update-alternatives --install /etc/sqoop/conf sqoop-conf /etc/sqoop/* 50
/usr/sbin/update-alternatives --install /etc/sqoop/conf sqoop-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/sqoop/conf.dist 10
/usr/sbin/update-alternatives --install /etc/sqoop2/conf sqoop2-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/sqoop2/conf.dist 10
/usr/sbin/update-alternatives --install /etc/zookeeper/conf zookeeper-conf
/cldr/app/coolapp/opt/parcels/CDH*/etc/zookeeper/conf.dist 10
/usr/sbin/update-alternatives --install /usr/bin/accumulo accumulo
/cldr/app/coolapp/opt/parcels/ACCUMULO-1.6.0-1.cdh5.1.0.p0.51/bin/accumulo 10
/usr/sbin/update-alternatives --install /usr/bin/accumulo-tool accumulo-tool
/cldr/app/coolapp/opt/parcels/ACCUMULO-1.6.0-1.cdh5.1.0.p0.51/bin/accumulo-tool 10
/usr/sbin/update-alternatives --install /usr/bin/avro-tools avro-tools
/cldr/app/coolapp/opt/parcels/CDH*/bin/avro-tools 10
/usr/sbin/update-alternatives --install /usr/bin/beeline beeline
/cldr/app/coolapp/opt/parcels/CDH*/bin/beeline 10
/usr/sbin/update-alternatives --install /usr/bin/catalogd catalogd
/cldr/app/coolapp/opt/parcels/CDH*/bin/catalogd 10
/usr/sbin/update-alternatives --install /usr/bin/cli_mt cli_mt
/cldr/app/coolapp/opt/parcels/CDH*/bin/cli_mt 10
/usr/sbin/update-alternatives --install /usr/bin/cli_st cli_st
/cldr/app/coolapp/opt/parcels/CDH*/bin/cli_st 10
/usr/sbin/update-alternatives --install /usr/bin/flume-ng flume-ng
/cldr/app/coolapp/opt/parcels/CDH*/bin/flume-ng 10
/usr/sbin/update-alternatives --install /usr/bin/hadoop hadoop
/cldr/app/coolapp/opt/parcels/CDH*/bin/hadoop 10
/usr/sbin/update-alternatives --install /usr/bin/hadoop-0.20 hadoop-0.20
/cldr/app/coolapp/opt/parcels/CDH*/bin/hadoop-0.20 10
/usr/sbin/update-alternatives --install /usr/bin/hadoop-fuse-dfs hadoop-fuse-dfs
/cldr/app/coolapp/opt/parcels/CDH*/bin/hadoop-fuse-dfs 10
/usr/sbin/update-alternatives --install /usr/bin/hbase hbase
/cldr/app/coolapp/opt/parcels/CDH*/bin/hbase 10
/usr/sbin/update-alternatives --install /usr/bin/hbase-indexer hbase-indexer
/cldr/app/coolapp/opt/parcels/CDH*/bin/hbase-indexer 10
/usr/sbin/update-alternatives --install /usr/bin/hcat hcat
/cldr/app/coolapp/opt/parcels/CDH*/bin/hcat 10
/usr/sbin/update-alternatives --install /usr/bin/hdfs hdfs
/cldr/app/coolapp/opt/parcels/CDH*/bin/hdfs 10
/usr/sbin/update-alternatives --install /usr/bin/hive hive
/cldr/app/coolapp/opt/parcels/CDH*/bin/hive 10
/usr/sbin/update-alternatives --install /usr/bin/hiveserver2 hiveserver2
/cldr/app/coolapp/opt/parcels/CDH*/bin/hiveserver2 10
/usr/sbin/update-alternatives --install /usr/bin/impala-shell impala-shell
/cldr/app/coolapp/opt/parcels/CDH*/bin/impala-shell 10
/usr/sbin/update-alternatives --install /usr/bin/impalad impalad
/cldr/app/coolapp/opt/parcels/CDH*/bin/impalad 10
/usr/sbin/update-alternatives --install /usr/bin/kite-dataset kite-dataset
/cldr/app/coolapp/opt/parcels/CDH*/bin/kite-dataset 10
/usr/sbin/update-alternatives --install /usr/bin/llama llama
/cldr/app/coolapp/opt/parcels/CDH*/bin/llama 10
/usr/sbin/update-alternatives --install /usr/bin/llamaadmin llamaadmin
/cldr/app/coolapp/opt/parcels/CDH*/bin/llamaadmin 10
/usr/sbin/update-alternatives --install /usr/bin/load_gen load_gen
/cldr/app/coolapp/opt/parcels/CDH*/bin/load_gen 10
/usr/sbin/update-alternatives --install /usr/bin/mahout mahout
/cldr/app/coolapp/opt/parcels/CDH*/bin/mahout 10
/usr/sbin/update-alternatives --install /usr/bin/mapred mapred
/cldr/app/coolapp/opt/parcels/CDH*/bin/mapred 10
/usr/sbin/update-alternatives --install /usr/bin/oozie oozie
/cldr/app/coolapp/opt/parcels/CDH*/bin/oozie 10
/usr/sbin/update-alternatives --install /usr/bin/pig pig
/cldr/app/coolapp/opt/parcels/CDH*/bin/pig 10
/usr/sbin/update-alternatives --install /usr/bin/pyspark pyspark
/cldr/app/coolapp/opt/parcels/CDH*/bin/pyspark 10
/usr/sbin/update-alternatives --install /usr/bin/sentry sentry
/cldr/app/coolapp/opt/parcels/CDH*/bin/sentry 10
/usr/sbin/update-alternatives --install /usr/bin/solrctl solrctl
```

```
/cldr/app/coolapp/opt/parcels/CDH*/bin/solrctl 10
/usr/sbin/update-alternatives --install /usr/bin/spark-executor spark-executor
/cldr/app/coolapp/opt/parcels/CDH*/bin/spark-executor 10
/usr/sbin/update-alternatives --install /usr/bin/spark-shell spark-shell
/cldr/app/coolapp/opt/parcels/CDH*/bin/spark-shell 10
/usr/sbin/update-alternatives --install /usr/bin/spark-submit spark-submit
/cldr/app/coolapp/opt/parcels/CDH*/bin/spark-submit 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop sqoop
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-codegen sqoop-codegen
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-codegen 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-create-hive-table
sqoop-create-hive-table /cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-create-hive-table
 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-eval sqoop-eval
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-eval 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-export sqoop-export
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-export 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-help sqoop-help
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-help 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-import sqoop-import
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-import 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-import-all-tables
sqoop-import-all-tables /cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-import-all-tables
 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-job sqoop-job
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-job 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-list-databases sqoop-list-databases
 /cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-list-databases 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-list-tables sqoop-list-tables
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-list-tables 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-merge sqoop-merge
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-merge 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-metastore sqoop-metastore
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-metastore 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop-version sqoop-version
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop-version 10
/usr/sbin/update-alternatives --install /usr/bin/sqoop2 sqoop2
/cldr/app/coolapp/opt/parcels/CDH*/bin/sqoop2 10
/usr/sbin/update-alternatives --install /usr/bin/statestored statestored
/cldr/app/coolapp/opt/parcels/CDH*/bin/statestored 10
/usr/sbin/update-alternatives --install /usr/bin/whirr whirr
/cldr/app/coolapp/opt/parcels/CDH*/bin/whirr 10
/usr/sbin/update-alternatives --install /usr/bin/yarn yarn
/cldr/app/coolapp/opt/parcels/CDH*/bin/yarn 10
/usr/sbin/update-alternatives --install /usr/bin/zookeeper-client zookeeper-client
/cldr/app/coolapp/opt/parcels/CDH*/bin/zookeeper-client 10
/usr/sbin/update-alternatives --install /usr/bin/zookeeper-server zookeeper-server
/cldr/app/coolapp/opt/parcels/CDH*/bin/zookeeper-server 10
/usr/sbin/update-alternatives --install /usr/bin/zookeeper-server-cleanup
zookeeper-server-cleanup /cldr/app/coolapp/opt/parcels/CDH*/bin/zookeeper-server-cleanup
 10
/usr/sbin/update-alternatives --install /usr/bin/zookeeper-server-initialize
zookeeper-server-initialize
/cldr/app/coolapp/opt/parcels/CDH*/bin/zookeeper-server-initialize 10
/usr/sbin/update-alternatives --install /usr/bin/bigtop-detect-javahome
bigtop-detect-javahome /cldr/app/coolapp/opt/parcels/CDH*/bin/bigtop-detect-javahome 10
/usr/sbin/update-alternatives --install /usr/bin/impala-collect-minidumps
impala-collect-minidumps /cldr/app/coolapp/opt/parcels/CDH*/bin impala-collect-minidumps
 10
/usr/sbin/update-alternatives --install /usr/bin/parquet-tools parquet-tools
/cldr/app/coolapp/opt/parcels/CDH*/bin/parquet-tools 10
/usr/sbin/update-alternatives --remove accumulo-conf /etc/accumulo/*
/usr/sbin/update-alternatives --remove hadoop-conf /etc/hadoop/*
/usr/sbin/update-alternatives --remove hbase-conf /etc/hbase/*
/usr/sbin/update-alternatives --remove hive-conf /etc/hive/*
/usr/sbin/update-alternatives --remove solr-conf /etc/solr/*
/usr/sbin/update-alternatives --remove spark-conf /etc/spark/*
/usr/sbin/update-alternatives --remove sqoop-conf /etc/sqoop/*
```

Commands run by the Cloudera Manager Agent while creating a diagnostic bundle:

```
/bin/bash -c cd ..; find -maxdepth
/bin/bash -c for x in /etc/security/limits.d/*;
/bin/bash -c PATH
/bin/cat /etc/apt/sources.list
/bin/cat /etc/host.conf
/bin/cat /etc/hostname
/bin/cat /etc/hosts
/bin/cat /etc/issue
/bin/cat /etc/krb5.conf
/bin/cat /etc/nsswitch.conf
/bin/cat /etc/redhat-release
/bin/cat /etc/resolv.conf
/bin/cat /etc/security/limits.conf
/bin/cat /etc/suse-release
/bin/cat /etc/sysconfig/network
/bin/cat /etc/sysconfig/network/ifcfg-eth0
/bin/cat /etc/sysconfig/network-scripts/ifcfg-eth0
/bin/cat /etc/sysconfig/selinux
/bin/cat /proc/cpuinfo
/bin/cat /proc/diskstats
/bin/cat /proc/interrupts
/bin/cat /proc/meminfo
/bin/cat /proc/mounts
/bin/cat /proc/partitions
/bin/cat /proc/swaps
/bin/cat /proc/sys/vm/swappiness
/bin/cat /proc/vmstat
/bin/cat /var/kerberos/krb5kdc/kadm5.acl
/bin/cat /var/kerberos/krb5kdc/kdc.conf
/bin/cat /var/log/kern.log
/bin/cat /var/log/messages
/bin/date
/bin/df -i
/bin/df -k
/bin/dmesg
/bin/grep -r . /sys/kernel/mm
/bin/hostname --fqdn
/bin/ls /etc/yum.repos.d
/bin/netstat -s
/bin/ps aux
/bin/rpm -qa
/bin/uname -a
/bin/uname -r
/sbin/chkconfig --list
/sbin/ethtool eth0
/sbin/ethtool -S eth0
/sbin/ifconfig -a
/sbin/iptables -L -v -n
/sbin/lsmod
/sbin/lspci
/sbin/lvdisplay
/sbin/service --status-all
/sbin/sysctl -A
/usr/bin/curl -m 1 http://169.254.169.254/2011-01-01/meta-data/instance-type
/usr/bin/dig any test-1.ent.cloudera.com
/usr/bin/host -v -t A test-1.ent.cloudera.com
/usr/bin/lsb_release -a
/usr/bin/lscpu
/usr/bin/nslookup -query
/usr/bin/ntpstat
/usr/bin/python -c import socket; print socket.getfqdn();
/usr/bin/sar -A
/usr/bin/top -b -n 1
/usr/bin/uptime
/usr/bin/vmstat
/usr/sbin/dmidecode
/usr/sbin/lsof -n -P
/usr/sbin/ntpq -pn
```

Installing Cloudera Manager, CDH, and Managed Services

This procedure is recommended for installing Cloudera Manager and CDH for production environments. For a non-production "easy install," see Installing a Proof-of-Concept Cluster.

Before you begin the installation, make sure you have reviewed the requirements and other considerations described in Before You Install on page 9.

The general steps in the installation procedure are as follows:

Step 1: Configure a Repository for Cloudera Manager

Cloudera Manager is installed using package management tools such as `yum` for RHEL compatible systems, `zypper` for SLES, and `apt-get` for Ubuntu. These tools depend on access to repositories to install software. Cloudera maintains Internet-accessible repositories for CDH and Cloudera Manager installation files. You can also create your own internal repository for hosts that do not have Internet access.

> **Important:** The procedures in this topic always install the *latest* version of Cloudera Manager. If you need to install a specific, lower version of Cloudera Manager, you must download the contents of the repository and create an internal package repository to use for your installation.

For more information on creating an internal repository for Cloudera Manager, see Using an Internal Package Repository on page 48.

> **Important:** As of February 1, 2021, all downloads of CDH and Cloudera Manager require a username and password and use a modified URL. You must use the modified URL, including the username and password when downloading the repository contents described below. You may need to upgrade Cloudera Manager to a newer version that uses the modified URLs.
>
> This can affect new installations, upgrades, adding new hosts to a cluster, downloading a new parcel, and adding a new cluster.
>
> For more information, see Updating an existing CDH/Cloudera Manager deployment to access downloads with authentication.

To use the Cloudera repository:

RHEL compatible

1. Download the `cloudera-manager.repo` file for your OS version to the `/etc/yum.repos.d/` directory on the Cloudera Manager Server host.

   See the **Repo File** column in the Cloudera Manager Version and Download Information table for the URL.

   For example:

```
sudo wget <repo_file_url> -P /etc/yum.repos.d/
```

2.
3. Import the repository signing GPG key:

   - RHEL 7 compatible:

```
sudo rpm --import
https://username:password@archive.cloudera.com/p/cm5/redhat/7/x86_64/cm/RPM-GPG-KEY-cloudera
```

- RHEL 6 compatible:

```
sudo rpm --import
https://username:password@archive.cloudera.com/p/cm5/redhat/6/x86_64/cm/RPM-GPG-KEY-cloudera
```

- RHEL 5 compatible:

```
sudo rpm --import
https://username:password@archive.cloudera.com/p/cm5/redhat/5/x86_64/cm/RPM-GPG-KEY-cloudera
```

4. Edit the `cloudera-manager.repo` file in a text editor and add the following lines:

```
username=MyUserName
password=MyPassword
```

For example:

```
[cloudera-manager]
# Packages for Cloudera Manager, Version 5, on RedHat or CentOS 7 x86_64
name=Cloudera Manager
baseurl=https://archive.cloudera.com/p/cm5/redhat/7/x86_64/cm/5/
gpgkey =https://archive.cloudera.com/p/cm5/redhat/7/x86_64/cm/RPM-GPG-KEY-cloudera
gpgcheck = 1

username=MyUserName
password=MyPassword
```

5. Continue to Step 2: Install Java Development Kit on page 84.

SLES

1. Update your system package index by running:

```
sudo zypper refresh
```

2. Add the repo using `zypper addrepo`.

   See the **Repo File** column in the Cloudera Manager Version and Download Information table for the URL.

   For example:

```
sudo zypper addrepo -f
https://username:password@archive.cloudera.com/p/cm5/sles/12/x86_64/cm/cloudera-cm.repo
```

3. Edit the `/etc/zypp/repos.d/cloudera-cm.repo` file and add the following lines:

```
username=MyUserName
password=MyPassword
```

For example:

```
[cloudera-cm]
# Packages for Cloudera's Distribution for cm, Version 5, on SLES 11 x86_64
name=Cloudera's Distribution for cm, Version 5
baseurl=https://archive.cloudera.com/cm5/sles/12/x86_64/cm/5/
gpgkey = https://archive.cloudera.com/cm5/sles/12/x86_64/cm/RPM-GPG-KEY-cloudera
gpgcheck = 1

username=MyUserName
password=MyPassword
```

4. Import the repository signing GPG key:

```
sudo rpm --import
https://username:password@archive.cloudera.com/p/cm5/sles/12/x86_64/cm/RPM-GPG-KEY-cloudera
```

5. Continue to Step 2: Install Java Development Kit on page 84.

Ubuntu, Debian

1. Download the `cloudera.list` file for your OS version to the `/etc/apt/sources.list.d/` directory on the Cloudera Manager Server host.

   See the **Repo File** column in the Cloudera Manager Version and Download Information table for the URL.

2. Edit the `/etc/apt/sources.list.d/cloudera.list` file and add the following lines:

```
username=MyUserName
password=MyPassword
```

   For example:

```
# Packages for cm5, on Ubuntu 16.04 amd64
deb [arch=amd64] https://archive.cloudera.com/cm5/ubuntu/xenial/amd64/cm xenial-cm5
contrib
deb-src https://archive.cloudera.com/cm5/ubuntu/xenial/amd64/cm xenial-cm5 contrib

username=MyUserName
password=MyPassword
```

3. Import the repository signing GPG key. For example:

```
wget
https://username:password@archive.cloudera.com/p/cm5/ubuntu/xenial/amd64/cm/archive.key
sudo apt-key add archive.key
```

4. Update your system package index by running:

```
sudo apt-get update
```

5. Continue to Step 2: Install Java Development Kit on page 84.

Step 2: Install Java Development Kit

For the JDK, you can either install the Oracle JDK version provided by Cloudera using Cloudera Manager, a different Oracle JDK directly from Oracle, or OpenJDK. Most Linux distributions supported by Cloudera include OpenJDK, but manual installation instructions are provided below if needed.

OpenJDK is supported with Cloudera Enterprise 5.16.0 and higher.

Requirements

- The JDK must be 64-bit. Do not use a 32-bit JDK.
- The installed JDK must be a supported version as documented in CDH and Cloudera Manager Supported JDK Versions.
- The *same version* of the JDK must be installed on each cluster host.
- The JDK must be installed at `/usr/java/jdk-version`.

Important:

- The RHEL-compatible and Ubuntu operating systems supported by Cloudera Enterprise all use AES-256 encryption by default for tickets. To support AES-256 bit encryption in JDK versions lower than 1.8u161, you must install the Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy File on all cluster and Hadoop user machines. Cloudera Manager can automatically install the policy files, or you can install them manually. For JCE Policy File installation instructions, see the `README.txt` file included in the `jce_policy-x.zip` file. JDK 1.8u161 and higher enable unlimited strength encryption by default, and do not require policy files.
- On SLES platforms, do not install or try to use the IBM Java version bundled with the SLES distribution. CDH does not run correctly with that version.

Installing Oracle JDK Using Cloudera Manager

Note: Cloudera, Inc. acquired Oracle JDK software under the Oracle Binary Code License Agreement. Pursuant to Item D(v)(a) of the SUPPLEMENTAL LICENSE TERMS of the Oracle Binary Code License Agreement, use of JDK software is governed by the terms of the Oracle Binary Code License Agreement. By installing the JDK software, you agree to be bound by these terms. If you do not wish to be bound by these terms, then do not install the Oracle JDK.

After completing Step 1: Configure a Repository for Cloudera Manager on page 82, you can install the Oracle JDK on the Cloudera Manager Server host using `yum` as follows:

```
sudo yum install oracle-j2sdk1.7
```

You can use Cloudera Manager to install the JDK on the remaining cluster hosts in an upcoming step. Continue to Step 3: Install Cloudera Manager Server on page 86.

Manually Installing Oracle JDK

The Oracle JDK installer is available both as an RPM-based installer for RPM-based systems, and as a `.tar.gz` file. These instructions are for the `.tar.gz` file.

1. Download the `.tar.gz` file for one of the 64-bit supported versions of the Oracle JDK from Java SE 8 Downloads. (This link is correct at the time of writing, but can change.)

   Note: If you want to download the JDK directly using a utility such as `wget`, you must accept the Oracle license by configuring headers, which are updated frequently. Blog posts and Q&A sites can be a good source of information on how to download a particular JDK version using `wget`.

2. Extract the JDK to `/usr/java/`*`jdk-version`*. For example:

```
tar xvfz /path/to/jdk-8u<update_version>-linux-x64.tar.gz -C /usr/java/
```

3. Repeat this procedure on all cluster hosts. After you have finished, continue to Step 3: Install Cloudera Manager Server on page 86.

Manually Installing OpenJDK

Before installing Cloudera Manager and CDH, perform the steps in this section to install OpenJDK on all hosts in your cluster(s).

When you install Cloudera Enterprise, Cloudera Manager includes an option to install Oracle JDK. De-select this option.

OpenJDK is supported only with Cloudera Manager and CDH 5.16.

Log in to each host and run the following command:

RHEL

```
su -c yum install java-1.8.0-openjdk-devel
```

Ubuntu

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

Debian 8

```
apt-get install -t jessie-backports openjdk-8-jdk
```

SLES

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

Step 3: Install Cloudera Manager Server

In this step you install the Cloudera Manager Server packages on the Cloudera Manager host.

Install the Cloudera Manager Server Packages

1. Install the Cloudera Manager Server packages either on the host where the database is installed, or on a host that has access to the database. This host need not be a host in the cluster that you want to manage with Cloudera Manager. On the Cloudera Manager Server host, type the following commands to install the Cloudera Manager packages.

OS	Command
RHEL, if you have a yum repo configured	`sudo yum install cloudera-manager-daemons cloudera-manager-server`
RHEL,if you're manually transferring RPMs	`sudo yum --nogpgcheck localinstall cloudera-manager-daemons-*.rpm` `sudo yum --nogpgcheck localinstall cloudera-manager-server-*.rpm`
SLES	`sudo zypper install cloudera-manager-daemons cloudera-manager-server`
Ubuntu or Debian	`sudo apt-get install cloudera-manager-daemons cloudera-manager-server`

2. If you are using an Oracle database for Cloudera Manager, increase the Java heap size for Cloudera Manager Server. Edit the `/etc/default/cloudera-scm-server` file on the Cloudera Manager server host. Locate the line that begins with `export CM_JAVA_OPTS` and change the `-Xmx2G` option to `-Xmx4G`.
3. On all hosts, run the following command to install the Cloudera Manager agent:

OS	Command
RHEL, CentOS, Oracle Linux	`sudo yum install cloudera-manager-daemons cloudera-manager-agent`
SLES	`sudo zypper install cloudera-manager-daemons cloudera-manager-agent`
Ubuntu	`sudo apt-get install cloudera-manager-daemons cloudera-manager-agent`

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

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.

   For more information on Agent configuration options, see Agent Configuration File.

5. Start the Agents by running the following command on all hosts:

   RHEL 7, SLES 12, Debian 8, Ubuntu 16.04

```
sudo systemctl start cloudera-scm-agent
```

   If the agent starts without errors, no response displays.

   RHEL 5 or 6, SLES 11, Debian 6 or 7, Ubuntu 12.04, 14.04

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

   You should see the following:

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

Install and Configure Databases

After installing the Cloudera Manager Server packages, continue to Step 4: Install and Configure Databases on page 87.

Step 4: Install and Configure Databases

Cloudera Manager uses various databases and datastores to store information about the Cloudera Manager configuration, as well as information such as the health of the system, or task progress.

Although you can deploy different types of databases in a single environment, doing so can create unexpected complications. Cloudera recommends choosing one supported database provider for all of the Cloudera databases.

Cloudera recommends installing the databases on different hosts than the services. Separating databases from services can help isolate the potential impact from failure or resource contention in one or the other. It can also simplify management in organizations that have dedicated database administrators.

You can use your own PostgreSQL, MariaDB, MySQL, or Oracle database for the Cloudera Manager Server and other services that use databases. For information about planning, managing, and backing up Cloudera Manager data stores, see Storage Space Planning for Cloudera Manager on page 9 and Backing Up Databases.

Required Databases

The following components all require databases: Cloudera Manager Server, Oozie Server, Sqoop Server, Activity Monitor, Reports Manager, Hive Metastore Server, Hue Server, Sentry Server, Cloudera Navigator Audit Server, and Cloudera Navigator Metadata Server. The type of data contained in the databases and their relative sizes are as follows:

- Cloudera Manager Server - Contains all the information about services you have configured and their role assignments, all configuration history, commands, users, and running processes. This relatively small database (< 100 MB) is the most important to back up.

  Important: When you restart processes, the configuration for each of the services is redeployed using information saved in the Cloudera Manager database. If this information is not available, your cluster cannot start or function correctly. You must schedule and maintain regular backups of the Cloudera Manager database to recover the cluster in the event of the loss of this database. For more information, see Backing Up Databases.

- Oozie Server - Contains Oozie workflow, coordinator, and bundle data. Can grow very large.
- Sqoop Server - Contains entities such as the connector, driver, links and jobs. Relatively small.
- Activity Monitor - Contains information about past activities. In large clusters, this database can grow large. Configuring an Activity Monitor database is only necessary if a MapReduce service is deployed.
- Reports Manager - Tracks disk utilization and processing activities over time. Medium-sized.
- Hive Metastore Server - Contains Hive metadata. Relatively small.
- Hue Server - Contains user account information, job submissions, and Hive queries. Relatively small.
- Sentry Server - Contains authorization metadata. Relatively small.
- Cloudera Navigator Audit Server - Contains auditing information. In large clusters, this database can grow large.
- Cloudera Navigator Metadata Server - Contains authorization, policies, and audit report metadata. Relatively small.

The Host Monitor and Service Monitor services use local disk-based datastores. For more information, see Data Storage for Monitoring Data.

The JDBC connector for your database *must* be installed on the hosts where you assign the Activity Monitor and Reports Manager roles.

Installing and Configuring Databases

For instructions on installing and configuring databases for Cloudera Manager, CDH, and other managed services, see the instructions for the type of database you want to use:

Install and Configure MariaDB for Cloudera Software

To use a MariaDB database, follow these procedures. For information on compatible versions of MariaDB, see CDH and Cloudera Manager Supported Databases.

Installing MariaDB Server

Note:

- If you already have a MariaDB database set up, you can skip to the section Configuring and Starting the MariaDB Server on page 89 to verify that your MariaDB configurations meet the requirements for Cloudera Manager.
- It is important that the `datadir` directory (`/var/lib/mysql` by default), is on a partition that has sufficient free space. For more information, see Storage Space Planning for Cloudera Manager on page 9.

1. Install MariaDB server:

OS	Command
RHEL compatible	`sudo yum install mariadb-server`
SLES	`sudo zypper install mariadb-server` **Note:** Some SLES systems encounter errors when using the `zypper install` command. For more information on resolving this issue, see the Novell Knowledgebase topic, error running chkconfig.
Ubuntu	`sudo apt-get install mariadb-server`

If these commands do not work, you might need to add a repository or use a different `yum install` command, particularly on RHEL 6 compatible operating systems. For more assistance, see the following topics on the MariaDB website:

- **RHEL compatible:** Installing MariaDB with yum
- **SLES:** MariaDB Package Repository Setup and Usage
- **Ubuntu:** Installing MariaDB .deb Files

Configuring and Starting the MariaDB Server

Note: If you are making changes to an existing database, make sure to stop any services that use the database before continuing.

1. Stop the MariaDB server if it is running:
   - **RHEL 7 compatible:**

```
sudo systemctl stop mariadb
```

   - **RHEL 6 compatible, Ubuntu, SLES:**

```
sudo service mariadb stop
```

2. If they exist, move old InnoDB log files `/var/lib/mysql/ib_logfile0` and `/var/lib/mysql/ib_logfile1` out of `/var/lib/mysql/` to a backup location.
3. Determine the location of the option file, `my.cnf` (`/etc/my.cnf` by default).
4. Update `my.cnf` so that it conforms to the following requirements:
   - To prevent deadlocks, set the isolation level to `READ-COMMITTED`.
   - The default settings in the MariaDB installations in most distributions use conservative buffer sizes and memory usage. Cloudera Management Service roles need high write throughput because they might insert many records in the database. Cloudera recommends that you set the `innodb_flush_method` property to `O_DIRECT`.
   - Set the `max_connections` property according to the size of your cluster:
     - Fewer than 50 hosts - You can store more than one database (for example, both the Activity Monitor and Service Monitor) on the same host. If you do this, you should:
       - Put each database on its own storage volume.
       - 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.

  - More than 50 hosts - Do not store more than one database on the same host. Use a separate host for each database/host pair. The hosts do not need to be reserved exclusively for databases, but each database should be on a separate host.

- If the cluster has more than 1000 hosts, set the `max_allowed_packet` property to `16M`. Without this setting, the cluster may fail to start due to the following exception: `com.mysql.jdbc.PacketTooBigException`.
- Although binary logging is not a requirement for Cloudera Manager installations, it provides benefits such as MariaDB replication or point-in-time incremental recovery after a database restore. The provided example configuration enables the binary log. For more information, see The Binary Log.

Here is an option file with Cloudera recommended settings:

```
[mysqld]
datadir=/var/lib/mysql
socket=/var/lib/mysql/mysql.sock
transaction-isolation = READ-COMMITTED
# Disabling symbolic-links is recommended to prevent assorted security risks;
# to do so, uncomment this line:
symbolic-links = 0
# Settings user and group are ignored when systemd is used.
# If you need to run mysqld under a different user or group,
# customize your systemd unit file for mariadb according to the
# instructions in http://fedoraproject.org/wiki/Systemd

key_buffer = 16M
key_buffer_size = 32M
max_allowed_packet = 32M
thread_stack = 256K
thread_cache_size = 64
query_cache_limit = 8M
query_cache_size = 64M
query_cache_type = 1

max_connections = 550
#expire_logs_days = 10
#max_binlog_size = 100M

#log_bin should be on a disk with enough free space.
#Replace '/var/lib/mysql/mysql_binary_log' with an appropriate path for your
#system and chown the specified folder to the mysql user.
log_bin=/var/lib/mysql/mysql_binary_log

#In later versions of MariaDB, if you enable the binary log and do not set
#a server_id, MariaDB will not start. The server_id must be unique within
#the replicating group.
server_id=1

binlog_format = mixed

read_buffer_size = 2M
read_rnd_buffer_size = 16M
sort_buffer_size = 8M
join_buffer_size = 8M

# InnoDB settings
innodb_file_per_table = 1
innodb_flush_log_at_trx_commit  = 2
innodb_log_buffer_size = 64M
innodb_buffer_pool_size = 4G
innodb_thread_concurrency = 8
innodb_flush_method = O_DIRECT
innodb_log_file_size = 512M

[mysqld_safe]
log-error=/var/log/mariadb/mariadb.log
pid-file=/var/run/mariadb/mariadb.pid

#
```

```
# include all files from the config directory
#
!includedir /etc/my.cnf.d
```

5. If AppArmor is running on the host where MariaDB is installed, you might need to configure AppArmor to allow MariaDB to write to the binary.
6. Ensure the MariaDB server starts at boot:

OS	Command
RHEL 7 compatible	`sudo systemctl enable mariadb`
RHEL 6 compatible	`sudo chkconfig mariadb on`
SLES	`sudo chkconfig --add mariadb`
Ubuntu	`sudo chkconfig mariadb on` **Note:** `chkconfig` may not be available on recent Ubuntu releases. You may need to use Upstart to configure MariaDB to start automatically when the system boots. For more information, see the Ubuntu documentation or the Upstart Cookbook.

7. Start the MariaDB server:
   - **RHEL 7 compatible:**

```
sudo systemctl start mariadb
```

   - **RHEL 6 compatible, Ubuntu, SLES:**

```
sudo service mariadb start
```

8. Run `/usr/bin/mysql_secure_installation` to set the MariaDB root password and other security-related settings. In a new installation, the `root` password is blank. Press the **Enter** key when you're prompted for the root password. For the rest of the prompts, enter the responses listed below in **bold**:

```
sudo /usr/bin/mysql_secure_installation
```

```
[...]
Enter current password for root (enter for none):
OK, successfully used password, moving on...
[...]
Set root password? [Y/n] Y
New password:
Re-enter new password:
[...]
Remove anonymous users? [Y/n] Y
[...]
Disallow root login remotely? [Y/n] N
[...]
Remove test database and access to it [Y/n] Y
[...]
Reload privilege tables now? [Y/n] Y
[...]
All done!  If you've completed all of the above steps, your MariaDB
installation should now be secure.
```

```
Thanks for using MariaDB!
```

Installing the MySQL JDBC Driver for MariaDB

The MariaDB JDBC driver is not supported. Follow the steps in this section to install and use the MySQL JDBC driver instead.

Install the JDBC driver on the Cloudera Manager Server host, as well as any other hosts running services that require database access. For more information on Cloudera software that uses databases, see Required Databases on page 88.

Cloudera recommends that you consolidate all roles that require databases on a limited number of hosts, and install the driver on those hosts. Locating all such roles on the same hosts is recommended but not required. Make sure to install the JDBC driver on each host running roles that access the database.

Note: Cloudera recommends using only version 5.1 of the JDBC driver.

OS	Command

RHEL	**Important:** Using the `yum install` command to install the MySQL driver package before installing a JDK installs OpenJDK, and then uses the Linux `alternatives` command to set the system JDK to be OpenJDK. If you intend to use an Oracle JDK, make sure that it is installed before installing the MySQL driver using `yum install`.
	Alternatively, use the following procedure to manually install the driver.
	1. Download the MySQL JDBC driver from http://www.mysql.com/downloads/connector/j/5.1.html (in `.tar.gz` format).
	2. Extract the JDBC driver JAR file from the downloaded file. For example:
	`tar zxvf mysql-connector-java-5.1.46.tar.gz`
	3. Copy the JDBC driver, renamed, to `/usr/share/java/`. If the target directory does not yet exist, create it. For example:
	`sudo mkdir -p /usr/share/java/`
	`cd mysql-connector-java-5.1.46`
	`sudo cp mysql-connector-java-5.1.46-bin.jar /usr/share/java/mysql-connector-java.jar`

SLES	`sudo zypper install mysql-connector-java`
Ubuntu or Debian	`sudo apt-get install libmysql-java`

Creating Databases for Cloudera Software

Create databases and service accounts for components that require databases:

- Cloudera Manager Server
- Cloudera Management Service roles:
  - Activity Monitor (if using the MapReduce service in a CDH 5 cluster)

  - Reports Manager

- Hue
- Each Hive metastore
- Sentry Server
- Cloudera Navigator Audit Server
- Cloudera Navigator Metadata Server
- Oozie

Record the values you enter for database names, usernames, and passwords. The Cloudera Manager installation wizard requires this information to correctly connect to these databases.

1. Log in as the `root` user, or another user with privileges to create database and grant privileges:

```
mysql -u root -p
```

```
Enter password:
```

2. Create databases for each service deployed in the cluster using the following commands. You can use any value you want for the *<database>*, *<user>*, and *<password>* parameters. The **Databases for Cloudera Software** table, below lists the default names provided in the Cloudera Manager configuration settings, but you are not required to use them.

   Configure all databases to use the `utf8` character set.

   Include the character set for each database when you run the `CREATE DATABASE` statements described below.

```
CREATE DATABASE <database> DEFAULT CHARACTER SET <character set> DEFAULT COLLATE
utf8_general_ci;
```

```
Query OK, 1 row affected (0.00 sec)
```

```
GRANT ALL ON <database>.* TO '<user>'@'%' IDENTIFIED BY '<password>';
```

```
Query OK, 0 rows affected (0.00 sec)
```

Table 14: Databases for Cloudera Software

Service	Database	User
Cloudera Manager Server	scm	scm
Activity Monitor	amon	amon
Reports Manager	rman	rman
Hue	hue	hue
Hive Metastore Server	metastore	hive
Sentry Server	sentry	sentry
Cloudera Navigator Audit Server	nav	nav
Cloudera Navigator Metadata Server	navms	navms
Oozie	oozie	oozie

3. Confirm that you have created all of the databases:

```
SHOW DATABASES;
```

You can also confirm the privilege grants for a given user by running:

```
SHOW GRANTS FOR '<user>'@'%';
```

Setting Up the Cloudera Manager Database

After completing the above instructions to install and configure MariaDB databases for Cloudera software, continue to Step 5: Set up the Cloudera Manager Database on page 115 to configure a database for Cloudera Manager.

Install and Configure MySQL for Cloudera Software

To use a MySQL database, follow these procedures. For information on compatible versions of the MySQL database, see CDH and Cloudera Manager Supported Databases.

Installing the MySQL Server

Note:

- If you already have a MySQL database set up, you can skip to the section Configuring and Starting the MySQL Server on page 95 to verify that your MySQL configurations meet the requirements for Cloudera Manager.
- For MySQL 5.6 and 5.7, you must install the *MySQL-shared-compat* or *MySQL-shared* package. This is required for the Cloudera Manager Agent package installation.
- It is important that the `datadir` directory, which, by default, is `/var/lib/mysql`, is on a partition that has sufficient free space.
- Cloudera Manager installation fails if GTID-based replication is enabled in MySQL.
- For Cloudera Navigator, make sure that the MySQL server system variable `explicit_defaults_for_timestamp` is disabled (set to "0") during installation and upgrades. (MySQL 5.6.6 and later).

1. Install the MySQL database.

OS	Command

RHEL	MySQL is no longer included with RHEL. You must download the repository from the MySQL site and install it directly. You can use the following commands to install MySQL. For more information, visit the MySQL website.
	`wget http://repo.mysql.com/mysql-community-release-el7-5.noarch.rpm`
	`sudo rpm -ivh mysql-community-release-el7-5.noarch.rpm`
	`sudo yum update`
	`sudo yum install mysql-server`
	`sudo systemctl start mysqld`

OS	Command
SLES	`sudo zypper install mysql libmysqlclient_r17` **Note:** Some SLES systems encounter errors when using the preceding `zypper install` command. For more information on resolving this issue, see the Novell Knowledgebase topic, error running chkconfig.
Ubuntu	`sudo apt-get install mysql-server`

Configuring and Starting the MySQL Server

Note: If you are making changes to an existing database, make sure to stop any services that use the database before continuing.

1. Stop the MySQL server if it is running.

OS	Command
RHEL 7 Compatible	`sudo systemctl stop mysqld`
RHEL 6 Compatible	`sudo service mysqld stop`
SLES, Ubuntu	`sudo service mysql stop`

2. Move old InnoDB log files `/var/lib/mysql/ib_logfile0` and `/var/lib/mysql/ib_logfile1` out of `/var/lib/mysql/` to a backup location.
3. Determine the location of the option file, `my.cnf` (`/etc/my.cnf` by default).
4. Update `my.cnf` so that it conforms to the following requirements:
   - To prevent deadlocks, set the isolation level to `READ-COMMITTED`.
   - Configure the `InnoDB` engine. Cloudera Manager will not start if its tables are configured with the MyISAM engine. (Typically, tables revert to MyISAM if the InnoDB engine is misconfigured.) To check which engine your tables are using, run the following command from the MySQL shell:

```
mysql> show table status;
```

   - The default settings in the MySQL installations in most distributions use conservative buffer sizes and memory usage. Cloudera Management Service roles need high write throughput because they might insert many records in the database. Cloudera recommends that you set the `innodb_flush_method` property to `O_DIRECT`.
   - Set the `max_connections` property according to the size of your cluster:
     - Fewer than 50 hosts - You can store more than one database (for example, both the Activity Monitor and Service Monitor) on the same host. If you do this, you should:
       - Put each database on its own storage volume.
       - 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.

  - More than 50 hosts - Do not store more than one database on the same host. Use a separate host for each database/host pair. The hosts do not need to be reserved exclusively for databases, but each database should be on a separate host.

- If the cluster has more than 1000 hosts, set the `max_allowed_packet` property to `16M`. Without this setting, the cluster may fail to start due to the following exception: `com.mysql.jdbc.PacketTooBigException`.
- Binary logging is not a requirement for Cloudera Manager installations. Binary logging provides benefits such as MySQL replication or point-in-time incremental recovery after database restore. Examples of this configuration follow. For more information, see The Binary Log.

Here is an option file with Cloudera recommended settings:

```
[mysqld]
datadir=/var/lib/mysql
socket=/var/lib/mysql/mysql.sock
transaction-isolation = READ-COMMITTED
# Disabling symbolic-links is recommended to prevent assorted security risks;
# to do so, uncomment this line:
symbolic-links = 0

key_buffer_size = 32M
max_allowed_packet = 16M
thread_stack = 256K
thread_cache_size = 64
query_cache_limit = 8M
query_cache_size = 64M
query_cache_type = 1

max_connections = 550
#expire_logs_days = 10
#max_binlog_size = 100M

#log_bin should be on a disk with enough free space.
#Replace '/var/lib/mysql/mysql_binary_log' with an appropriate path for your
#system and chown the specified folder to the mysql user.
log_bin=/var/lib/mysql/mysql_binary_log

#In later versions of MySQL, if you enable the binary log and do not set
#a server_id, MySQL will not start. The server_id must be unique within
#the replicating group.
server_id=1

binlog_format = mixed

read_buffer_size = 2M
read_rnd_buffer_size = 16M
sort_buffer_size = 8M
join_buffer_size = 8M

# InnoDB settings
innodb_file_per_table = 1
innodb_flush_log_at_trx_commit  = 2
innodb_log_buffer_size = 64M
innodb_buffer_pool_size = 4G
innodb_thread_concurrency = 8
innodb_flush_method = O_DIRECT
innodb_log_file_size = 512M

[mysqld_safe]
log-error=/var/log/mysqld.log
pid-file=/var/run/mysqld/mysqld.pid

sql_mode=STRICT_ALL_TABLES
```

5. If AppArmor is running on the host where MySQL is installed, you might need to configure AppArmor to allow MySQL to write to the binary.
6. Ensure the MySQL server starts at boot:

OS	Command
RHEL 7 compatible	`sudo systemctl enable mysqld`
RHEL 6 compatible	`sudo chkconfig mysqld on`
SLES	`sudo chkconfig --add mysql`
Ubuntu	`sudo chkconfig mysql on` **Note:** `chkconfig` may not be available on recent Ubuntu releases. You may need to use Upstart to configure MySQL to start automatically when the system boots. For more information, see the Ubuntu documentation or the Upstart Cookbook.

7. Start the MySQL server:

OS	Command
RHEL 7 Compatible	`sudo systemctl start mysqld`
RHEL 6 Compatible	`sudo service mysqld start`
SLES, Ubuntu	`sudo service mysql start`

8. Run `/usr/bin/mysql_secure_installation` to set the MySQL root password and other security-related settings. In a new installation, the `root` password is blank. Press the **Enter** key when you're prompted for the root password. For the rest of the prompts, enter the responses listed below in **bold**:

```
sudo /usr/bin/mysql_secure_installation
```

```
[...]
Enter current password for root (enter for none):
OK, successfully used password, moving on...
[...]
Set root password? [Y/n] Y
New password:
Re-enter new password:
Remove anonymous users? [Y/n] Y
[...]
Disallow root login remotely? [Y/n] N
[...]
Remove test database and access to it [Y/n] Y
[...]
Reload privilege tables now? [Y/n] Y
All done!
```

Installing the MySQL JDBC Driver

Install the JDBC driver on the Cloudera Manager Server host, as well as any other hosts running services that require database access. For more information on Cloudera software that uses databases, see Required Databases on page 88.

Note: If you already have the JDBC driver installed on the hosts that need it, you can skip this section. However, MySQL 5.6 requires a 5.1 driver version 5.1.26 or higher.

Cloudera recommends that you consolidate all roles that require databases on a limited number of hosts, and install the driver on those hosts. Locating all such roles on the same hosts is recommended but not required. Make sure to install the JDBC driver on each host running roles that access the database.

Note: Cloudera recommends using only version 5.1 of the JDBC driver.

OS	Command

RHEL	**Important:** Using the `yum install` command to install the MySQL driver package before installing a JDK installs OpenJDK, and then uses the Linux `alternatives` command to set the system JDK to be OpenJDK. If you intend to use an Oracle JDK, make sure that it is installed before installing the MySQL driver using `yum install`.
	Alternatively, use the following procedure to manually install the driver.
	1. Download the MySQL JDBC driver from http://www.mysql.com/downloads/connector/j/5.1.html (in `.tar.gz` format).
	2. Extract the JDBC driver JAR file from the downloaded file. For example:
	`tar zxvf mysql-connector-java-5.1.46.tar.gz`
	3. Copy the JDBC driver, renamed, to `/usr/share/java/`. If the target directory does not yet exist, create it. For example:
	`sudo mkdir -p /usr/share/java/`
	`cd mysql-connector-java-5.1.46`
	`sudo cp mysql-connector-java-5.1.46-bin.jar /usr/share/java/mysql-connector-java.jar`

SLES	`sudo zypper install mysql-connector-java`
Ubuntu or Debian	`sudo apt-get install libmysql-java`

Creating Databases for Cloudera Software

Create databases and service accounts for components that require databases:

- Cloudera Manager Server
- Cloudera Management Service roles:
  - Activity Monitor (if using the MapReduce service in a CDH 5 cluster)
  - Reports Manager
- Hue
- Each Hive metastore
- Sentry Server
- Cloudera Navigator Audit Server
- Cloudera Navigator Metadata Server

- Oozie

Record the values you enter for database names, usernames, and passwords. The Cloudera Manager installation wizard requires this information to correctly connect to these databases.

1. Log in as the `root` user, or another user with privileges to create database and grant privileges:

```
mysql -u root -p
```

```
Enter password:
```

2. Create databases for each service deployed in the cluster using the following commands. You can use any value you want for the *<database>*, *<user>*, and *<password>* parameters. The **Databases for Cloudera Software** table, below lists the default names provided in the Cloudera Manager configuration settings, but you are not required to use them.

   Configure all databases to use the `utf8` character set.

   Include the character set for each database when you run the `CREATE DATABASE` statements described below.

```
CREATE DATABASE <database> DEFAULT CHARACTER SET <character set> DEFAULT COLLATE
utf8_general_ci;
```

```
Query OK, 1 row affected (0.00 sec)
```

```
GRANT ALL ON <database>.* TO '<user>'@'%' IDENTIFIED BY '<password>';
```

```
Query OK, 0 rows affected (0.00 sec)
```

Table 15: Databases for Cloudera Software

Service	Database	User
Cloudera Manager Server	scm	scm
Activity Monitor	amon	amon
Reports Manager	rman	rman
Hue	hue	hue
Hive Metastore Server	metastore	hive
Sentry Server	sentry	sentry
Cloudera Navigator Audit Server	nav	nav
Cloudera Navigator Metadata Server	navms	navms
Oozie	oozie	oozie

3. Confirm that you have created all of the databases:

```
SHOW DATABASES;
```

   You can also confirm the privilege grants for a given user by running:

```
SHOW GRANTS FOR '<user>'@'%';
```

Setting Up the Cloudera Manager Database

After completing the above instructions to install and configure MySQL databases for Cloudera software, continue to Step 5: Set up the Cloudera Manager Database on page 115 to configure a database for Cloudera Manager.

Install and Configure PostgreSQL for Cloudera Software

> **Note:** The following instructions are for a dedicated PostgreSQL database for use in production environments, and are unrelated to the embedded PostgreSQL database provided by Cloudera for non-production installations.

To use a PostgreSQL database, follow these procedures. For information on compatible versions of the PostgreSQL database, see CDH and Cloudera Manager Supported Databases.

Installing PostgreSQL Server

> **Note:**
> - If you already have a PostgreSQL database set up, you can skip to the section Configuring and Starting the PostgreSQL Server on page 101 to verify that your PostgreSQL configurations meet the requirements for Cloudera Manager.
> - Make sure that the data directory, which by default is `/var/lib/postgresql/data/`, is on a partition that has sufficient free space.
> - Cloudera Manager supports the use of a custom schema name for the Cloudera Manager Server database, but not the CDH component databases (such as Hive, Hue, Sentry, and so on). For more information, see https://www.postgresql.org/docs/current/static/ddl-schemas.html.

Install the PostgreSQL packages as follows:

RHEL:

```
sudo yum install postgresql-server
```

SLES:

```
sudo zypper install --no-recommends postgresql96-server
```

> **Note:** This command installs PostgreSQL 9.6. If you want to install a different version, you can use `zypper search postgresql` to search for an available supported version. See CDH and Cloudera Manager Supported Databases.

Ubuntu:

```
sudo apt-get install postgresql
```

Installing `psycopg2` Python Package

Hue relies on the `psycopg2` Python package for connecting to a PostgreSQL database. Install the `psycopg2` package as follows:

RHEL compatible:

1. Install the `python-pip` package:

```
sudo yum install python-pip
```

2. Install psycopg2 using pip:

```
sudo pip install psycopg2
```

3. Link the psycopg2 directory to the Hue Python environment:

```
sudo ln -s /usr/lib64/python2.7/site-packages/psycopg2
/opt/cloudera/parcels/CDH/lib/hue/build/env/lib/python2.7/site-packages/psycopg2
```

SLES:

1. Install the python-psycopg2 package:

```
sudo zypper install python-psycopg2
```

2. Link the psycopg2 directory to the Hue Python environment:

```
sudo ln -s /usr/lib64/python2.7/site-packages/psycopg2
/opt/cloudera/parcels/CDH/lib/hue/build/env/lib/python2.7/site-packages/psycopg2
```

Ubuntu:

1. Install the python-pip package:

```
sudo apt-get install python-pip
```

2. Install psycopg2 using pip:

```
sudo pip install psycopg2
```

3. Link the psycopg2 directory to the Hue Python environment:

```
sudo ln -s /usr/lib64/python2.7/site-packages/psycopg2
/opt/cloudera/parcels/CDH/lib/hue/build/env/lib/python2.7/site-packages/psycopg2
```

Configuring and Starting the PostgreSQL Server

Note: If you are making changes to an existing database, make sure to stop any services that use the database before continuing.

By default, PostgreSQL only accepts connections on the loopback interface. You must reconfigure PostgreSQL to accept connections from the fully qualified domain names (FQDN) of the hosts hosting the services for which you are configuring databases. If you do not make these changes, the services cannot connect to and use the database on which they depend.

1. Make sure that LC_ALL is set to en_US.UTF-8 and initialize the database as follows:

   - **RHEL 7:**

```
echo 'LC_ALL="en_US.UTF-8"' >> /etc/locale.conf
sudo su -l postgres -c "postgresql-setup initdb"
```

   - **RHEL 6:**

```
echo 'LC_ALL="en_US.UTF-8"' >> /etc/default/locale
sudo service postgresql initdb
```

- **SLES 12:**

```
sudo su -l postgres -c "initdb --pgdata=/var/lib/pgsql/data --encoding=UTF-8"
```

- **Ubuntu:**

```
sudo service postgresql start
```

2. Enable MD5 authentication. Edit `pg_hba.conf`, which is usually found in `/var/lib/pgsql/data` or `/etc/postgresql/<version>/main`. Add the following line:

```
host all all 127.0.0.1/32 md5
```

If the default `pg_hba.conf` file contains the following line:

```
host all all 127.0.0.1/32 ident
```

then the `host` line specifying `md5` authentication shown above must be inserted *before* this `ident` line. Failure to do so may cause an authentication error when running the `scm_prepare_database.sh` script. You can modify the contents of the `md5` line shown above to support different configurations. For example, if you want to access PostgreSQL from a different host, replace `127.0.0.1` with your IP address and update `postgresql.conf`, which is typically found in the same place as `pg_hba.conf`, to include:

```
listen_addresses = '*'
```

3. Configure settings to ensure your system performs as expected. Update these settings in the `/var/lib/pgsql/data/postgresql.conf` or `/var/lib/postgresql/data/postgresql.conf` file. Settings vary based on cluster size and resources as follows:

- Small to mid-sized clusters - Consider the following settings as starting points. If resources are limited, consider reducing the buffer sizes and checkpoint segments further. Ongoing tuning may be required based on each host's resource utilization. For example, if the Cloudera Manager Server is running on the same host as other roles, the following values may be acceptable:
  - `max_connection` - In general, allow each database on a host 100 maximum connections and then add 50 extra connections. You may have to increase the system resources available to PostgreSQL, as described at Connection Settings.
  - `shared_buffers` - 256MB
  - `wal_buffers` - 8MB
  - `checkpoint_segments` - 16

Note: The `checkpoint_segments` setting is removed in PostgreSQL 9.5 and higher, replaced by `min_wal_size` and `max_wal_size`. The PostgreSQL 9.5 release notes provides the following formula for determining the new settings:

```
max_wal_size = (3 * checkpoint_segments) * 16MB
```

  - `checkpoint_completion_target` - 0.9

- Large clusters - Can contain up to 1000 hosts. Consider the following settings as starting points.
  - `max_connection` - For large clusters, each database is typically hosted on a different host. In general, allow each database on a host 100 maximum connections and then add 50 extra connections. You may have to increase the system resources available to PostgreSQL, as described at Connection Settings.
  - `shared_buffers` - 1024 MB. This requires that the operating system can allocate sufficient shared memory. See PostgreSQL information on Managing Kernel Resources for more information on setting kernel resources.

- `wal_buffers` - 16 MB. This value is derived from the `shared_buffers` value. Setting `wal_buffers` to be approximately 3% of `shared_buffers` up to a maximum of approximately 16 MB is sufficient in most cases.
- `checkpoint_segments` - 128. The PostgreSQL Tuning Guide recommends values between 32 and 256 for write-intensive systems, such as this one.

> **Note:** The `checkpoint_segments` setting is removed in PostgreSQL 9.5 and higher, replaced by `min_wal_size` and `max_wal_size`. The PostgreSQL 9.5 release notes provides the following formula for determining the new settings:
>
> ```
> max_wal_size = (3 * checkpoint_segments) * 16MB
> ```

- `checkpoint_completion_target` - 0.9.

4. Configure the PostgreSQL server to start at boot.

OS	Command
RHEL 7 compatible	`sudo systemctl enable postgresql`
RHEL 6 compatible	`sudo chkconfig postgresql on`
SLES	`sudo chkconfig --add postgresql`
Ubuntu	`sudo chkconfig postgresql on` **Note:** `chkconfig` may not be available on recent Ubuntu releases. You may need to use Upstart to configure PostgreSQL to start automatically when the system boots. For more information, see the Ubuntu documentation or the Upstart Cookbook.

5. Restart the PostgreSQL database:
   - **RHEL 7 Compatible:**

```
sudo systemctl restart postgresql
```

   - **All Others:**

```
sudo service postgresql restart
```

Creating Databases for Cloudera Software

Create databases and service accounts for components that require databases:

- Cloudera Manager Server
- Cloudera Management Service roles:
  - Activity Monitor (if using the MapReduce service in a CDH 5 cluster)
  - Reports Manager
- Hue

- Each Hive metastore
- Sentry Server
- Cloudera Navigator Audit Server
- Cloudera Navigator Metadata Server
- Oozie

The databases must be configured to support the PostgreSQL UTF8 character set encoding.

Record the values you enter for database names, usernames, and passwords. The Cloudera Manager installation wizard requires this information to correctly connect to these databases.

1. Connect to PostgreSQL:

```
sudo -u postgres psql
```

2. Create databases for each service you are using from the below table:

```
CREATE ROLE <user> LOGIN PASSWORD '<password>';
```

```
CREATE DATABASE <database> OWNER <user> ENCODING 'UTF8';
```

You can use any value you want for *<database>*, *<user>*, and *<password>*. The following examples are the default names provided in the Cloudera Manager configuration settings, but you are not required to use them:

Table 16: Databases for Cloudera Software

Service	Database	User
Cloudera Manager Server	scm	scm
Activity Monitor	amon	amon
Reports Manager	rman	rman
Hue	hue	hue
Hive Metastore Server	metastore	hive
Sentry Server	sentry	sentry
Cloudera Navigator Audit Server	nav	nav
Cloudera Navigator Metadata Server	navms	navms
Oozie	oozie	oozie

Record the databases, usernames, and passwords chosen because you will need them later.

3. For PostgreSQL 8.4 and higher, set `standard_conforming_strings=off` for the Hive Metastore and Oozie databases:

```
ALTER DATABASE <database> SET standard_conforming_strings=off;
```

Setting Up the Cloudera Manager Database

After completing the above instructions to install and configure PostgreSQL databases for Cloudera software, continue to Step 5: Set up the Cloudera Manager Database on page 115 to configure a database for Cloudera Manager.

Install and Configure Oracle Database for Cloudera Software

To use an Oracle database, follow these procedures. For information on compatible versions of the Oracle database, see CDH and Cloudera Manager Supported Databases.

Collecting Oracle Database Information

To configure Cloudera Manager to work with an Oracle database, get the following information from your Oracle DBA:

- Hostname - The DNS name or the IP address of the host where the Oracle database is installed.
- SID - The name of the schema that will store Cloudera Manager information.
- Username - A username for each schema that is storing information. You could have four unique usernames for the four schema.
- Password - A password corresponding to each username.

Configuring the Oracle Server

Note: If you are making changes to an existing database, make sure to stop any services that use the database before continuing.

Adjusting Oracle Settings to Accommodate Larger Clusters

Cloudera Management services require high write throughput. Depending on the size of your deployments, your DBA may need to modify Oracle settings for monitoring services. These guidelines are for larger clusters and do not apply to the Cloudera Manager configuration database and to smaller clusters. Many factors help determine whether you need to change your database settings, but in most cases, if your cluster has more than 100 hosts, you should consider making the following changes:

- Enable direct and asynchronous I/O by setting the `FILESYSTEMIO_OPTIONS` parameter to `SETALL`.
- Increase the RAM available to Oracle by changing the `MEMORY_TARGET` parameter. The amount of memory to assign depends on the size of the Hadoop cluster.
- Create more redo log groups and spread the redo log members across separate disks or logical unit numbers.
- Increase the size of redo log members to be at least 1 GB.

Reserving Ports for HiveServer 2

HiveServer2 uses port 10000 by default, but Oracle database changes the local port range. This can cause HiveServer2 to fail to start.

Manually reserve the default port for HiveServer2. For example, the following command reserves port 10000 and inserts a comment indicating the reason:

```
echo << EOF > /etc/sysctl.cnf
# HS2 uses port 10000
net.ipv4.ip_local_reserved_ports = 10000
EOF
```

```
sysctl -q -w net.ipv4.ip_local_reserved_ports=10000
```

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.

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;
```

Ensuring Your Oracle Database Supports UTF8

The database you use must support UTF8 character set encoding. You can implement UTF8 character set encoding in Oracle databases by using the dbca utility. In this case, you can use the `characterSet AL32UTF8` option to specify proper encoding. Consult your DBA to ensure UTF8 encoding is properly configured.

Installing the Oracle JDBC Connector

You must install the JDBC connector on the Cloudera Manager Server host and any other hosts that use a database.

Cloudera recommends that you assign all roles that require a database on the same host and install the connector on that host. Locating all such roles on the same host is recommended but not required. If you install a role, such as Activity Monitor, on one host and other roles on a separate host, you would install the JDBC connector on each host running roles that access the database.

1. Download the Oracle JDBC Driver from the Oracle website. For example, the version 8 JAR file is named `ojdbc8.jar`.

   For more information about supported Oracle Java versions, see CDH and Cloudera Manager Supported JDK Versions.

   To download the JDBC driver, visit the Oracle JDBC and UCP Downloads page, and click on the link for your Oracle Database version. Download the `ojdbc7.jar` file (or `ojdbc8.jar`, which is required for Oracle 18 or 19).

2. Copy the Oracle JDBC JAR file to `/usr/share/java/oracle-connector-java.jar`. The Cloudera Manager databases and the Hive Mestastore database use this shared file. For example:

```
mkdir /usr/share/java
cp /tmp/ojdbc8.jar /usr/share/java/oracle-connector-java.jar
```

Creating Databases for Cloudera Software

Create schema and user accounts for components that require databases:

- Cloudera Manager Server
- Cloudera Management Service roles:
  - Activity Monitor (if using the MapReduce service in a CDH 5 cluster)
  - Reports Manager
- Hue
- Each Hive metastore
- Sentry Server
- Cloudera Navigator Audit Server

- Cloudera Navigator Metadata Server
- Oozie

You can create the Oracle database, schema and users on the host where the Cloudera Manager Server will run, or on any other hosts in the cluster. For performance reasons, you should install each database on the host on which the service runs, as determined by the roles you assign during installation or upgrade. In larger deployments or in cases where database administrators are managing the databases the services use, you can separate databases from services, but use caution.

The database must be configured to support UTF-8 character set encoding.

Record the values you enter for database names, usernames, and passwords. The Cloudera Manager installation wizard requires this information to correctly connect to these databases.

1. Log into the Oracle client:

```
sqlplus system@localhost
```

```
Enter password: ******
```

2. Create a user and schema for each service you are using from the below table:

```
create user <user> identified by <password> default tablespace <tablespace>;
grant CREATE SESSION to <user>;
grant CREATE TABLE to <user>;
grant CREATE SEQUENCE to <user>;
grant EXECUTE on sys.dbms_lob to <user>;
```

You can use any value you want for *<schema>*, *<user>*, and *<password>*. The following examples are the default names provided in the Cloudera Manager configuration settings, but you are not required to use them:

Table 17: Databases for Cloudera Software

Service	Database	User
Cloudera Manager Server	scm	scm
Activity Monitor	amon	amon
Reports Manager	rman	rman
Hue	hue	hue
Hive Metastore Server	metastore	hive
Sentry Server	sentry	sentry
Cloudera Navigator Audit Server	nav	nav
Cloudera Navigator Metadata Server	navms	navms
Oozie	oozie	oozie

3. Grant a quota on the tablespace (the default tablespace is SYSTEM) where tables will be created:

```
ALTER USER <user> quota 100m on <tablespace>;
```

or for unlimited space:

```
ALTER USER username quota unlimited on <tablespace>;
```

4. Set the following additional privileges for Oozie:

```
grant alter index to oozie;
grant alter table to oozie;
grant create index to oozie;
grant create sequence to oozie;
grant create session to oozie;
grant create table to oozie;
grant drop sequence to oozie;
grant select dictionary to oozie;
grant drop table to oozie;
alter user oozie quota unlimited on <tablespace>;
```

> **Important:**
>
> For security reasons, *do not* grant `select any table` privileges to the Oozie user.

5. Set the following additional privileges for the Cloudera Navigator Audit Server database:

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

where <*nav*> is the Navigator Audit Server user you specified above when you created the database.

For further information about Oracle privileges, see Authorization: Privileges, Roles, Profiles, and Resource Limitations.

Configuring the Hue Server to Store Data in Oracle (Client Parcel)

To install and configure the Oracle server and client repository for Hue, see Connect Hue to Oracle with Client Parcel

Connect Hue Service to Oracle

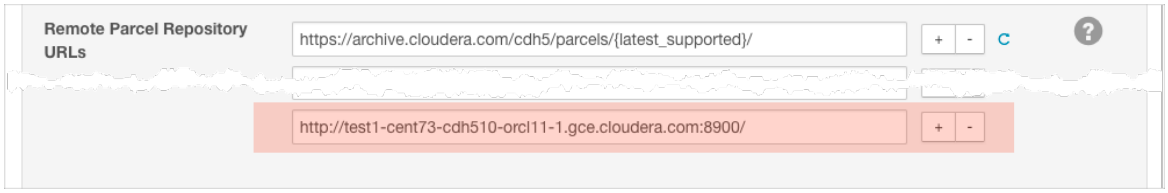
You can connect Hue to your Oracle database while installing CDH (and Hue) or with an existing installation. With existing CDH installations, you can connect and restart Hue, without saving the data in your current database, or you can migrate the old data into Oracle.

New CDH Installation

See Cloudera Installation Guide on page 8 to install Cloudera Manager (and its Installation Wizard), which you will use here to install CDH and the Oracle client.

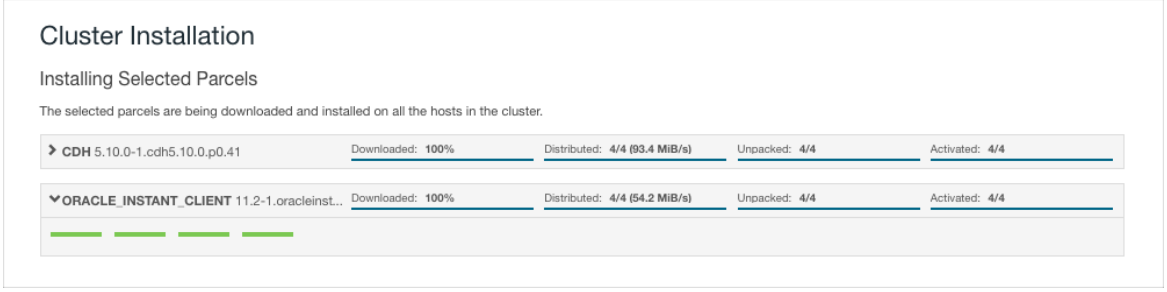
Install CDH and Oracle Parcel

1. Open the Cloudera Manager Admin Console and run the Cloudera Manager Installation Wizard to install CDH (and Hue). The URL for Cloudera Manager is: `http://<cm server hostname>:7180`
2. Stop at **Select Repository** to add the Oracle client parcel repository (**Cluster Installation**, step 1):
   a. Choose Method **Use Parcels** and click **More Options**.
   b. ➕:
   and add the URL for your Oracle **Remote Parcel Repository**:



   c. Click **Save Changes**.
   d. Select the newly added radio button by **ORACLE_INSTANT_CLIENT** and click **Continue**.

The Oracle parcel is downloaded, distributed, and activated at **Cluster Installation**, step 6 (**Installing Selected Parcels**).



Connect Hue to Oracle

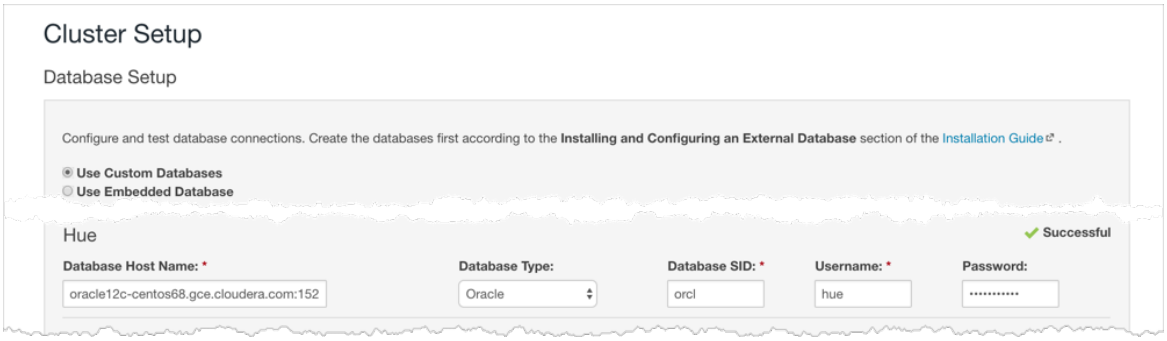
Continuing with Cloudera Manager Installation Wizard …

1. Stop at **Database Setup** to set connection properties (**Cluster Setup**, step 3).
   a. Select **Use Custom Database**.
   b. Under **Hue**, set the connection properties to the Oracle database.

> **Note:** Copy and store the password for the Hue embedded database (just in case).

```
Database Hostname (and port): <fqdn of host with Oracle server>:1521
Database Type (or engine): Oracle
Database SID (or name): orcl
Database Username: hue
Database Password: <hue database password>
```

   c. Click **Test Connection** and click **Continue** when successful.



2. Continue with the installation and click **Finish** to complete.
3. Add support for a multi-threaded environment:
   a. Go to **Clusters** > **Hue** > **Configuration**.

b. Filter by Category, **Hue-service** and Scope, **Advanced**.
c. Add support for a multi-threaded environment by setting **Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini**:

```
[desktop]
[[database]]
options={"threaded":true}
```

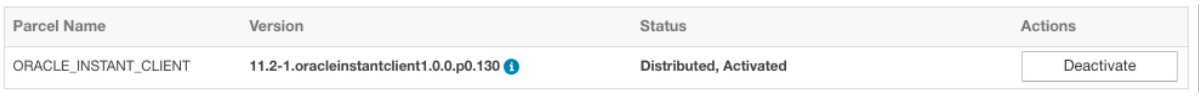
d. Click **Save Changes**.

4. Restart the Hue service: select **Actions** > **Restart** and click **Restart**.
5. Log on to Hue by clicking **Hue Web UI**.

Existing CDH Installation

Activate Oracle Client Parcel

1. Log on to Cloudera Manager.
2. Go to the **Parcels** page by clicking **Hosts** > **Parcels** (or clicking the parcels icon).
3. Click the **Configuration** > **Check for New Parcels**.
4. Find ORACLE_INSTANT_CLIENT and click **Download**, **Distribute**, and **Activate**.



Connect Hue to Oracle

If you are not migrating the current (or old) database, simply connect to your new Oracle database and restart Hue. See Connect Hue to an External Database.

1. [migration only] **Stop Hue Service**

a. In Cloudera Manager, navigate to **Cluster** > **Hue**.
b. Select **Actions** > **Stop**.

> **Note:** If necessary, refresh the page to ensure the Hue service is stopped: .

2. [migration only] **Dump Current Database**

a. Select **Actions** > **Dump Database**.
b. Click **Dump Database**. The file is written to `/tmp/hue_database_dump.json` on the host of the Hue server.
c. Log on to the *host of the Hue server* in a command-line terminal.
d. Edit **`/tmp/hue_database_dump.json`** by removing all objects with `useradmin.userprofile` in the `model` field. For example:

```
# Count number of objects
grep -c useradmin.userprofile /tmp/hue_database_dump.json
```

```
vi /tmp/hue_database_dump.json
```

```
{
  "pk": 1,
  "model": "useradmin.userprofile",
  "fields": {
    "last_activity": "2016-10-03T10:06:13",
    "creation_method": "HUE",
    "first_login": false,
    "user": 1,
    "home_directory": "/user/admin"
  }
```

```
},
{
  "pk": 2,
  "model": "useradmin.userprofile",
  "fields": {
    "last_activity": "2016-10-03T10:27:10",
    "creation_method": "HUE",
    "first_login": false,
    "user": 2,
    "home_directory": "/user/alice"
  }
},
```

3. **Connect to New Database**

   a. Configure Database connections:

   - Go to **Hue** > **Configuration** and filter by category, **Database**.
   - Set database properties and click **Save Changes**:

```
Hue Database Type (or engine): Oracle
Hue Database Hostname: <fqdn of host with Oracle server>
Hue Database Port: 1521
Hue Database Username: hue
Hue Database Password: <hue database password>
Hue Database Name (or SID): orcl
```

   b. Add support for a multi-threaded environment:

   - Filter by Category, **Hue-service** and Scope, **Advanced**.
   - Set **Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini** and click **Save Changes**:

```
[desktop]
[[database]]
options={"threaded":true}
```

4. [migration only] **Synchronize New Database**

   a. Select **Actions** > **Synchronize Database**
   b. Click **Synchronize Database**.

5. [migration only] **Load Data from Old Database**

> **Important:** All user tables in the Hue database must be empty. You cleaned them at step 3 of Create Hue Database. Ensure they are still clean.

```
sqlplus hue/<your hue password> < delete_from_tables.ddl
```

6. **Re/Start Hue service**

   a. Navigate to **Cluster** > **Hue**.
   b. Select **Actions** > **Start**, and click **Start**.
   c. Click **Hue Web UI** to log on to Hue with a custom Oracle database.

Configuring the Hue Server to Store Data in Oracle (Client Package)

To install and configure the Oracle server and client repository for Hue, see Connect Hue to Oracle with Client Package

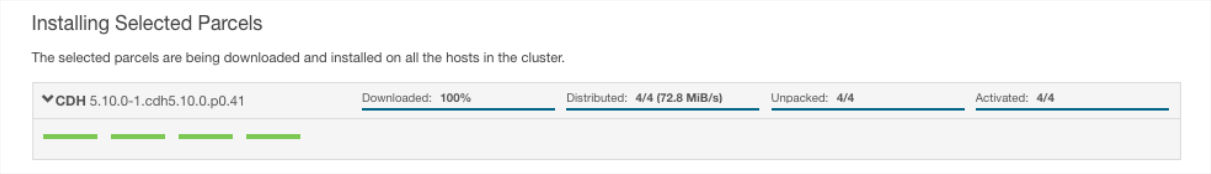
Connect Hue Service to Oracle

You can connect Hue to your Oracle database while installing CDH (and Hue) or with an existing installation. With existing CDH installations, you can connect and restart Hue, without saving the data in your current database, or you can migrate the old data into Oracle.

New CDH Installation

See Cloudera Installation Guide on page 8 to install Cloudera Manager (and its Installation Wizard), which you will use here to install CDH and the Oracle client.

1. Open the Cloudera Manager Admin Console and run the Cloudera Manager Installation Wizard to install CDH (and Hue). The URL for Cloudera Manager is: `http://<cm server hostname>:7180`
2. Stop at the end of **Cluster Installation** to copy the latest `cx_Oracle` package into Hue's Python environment.

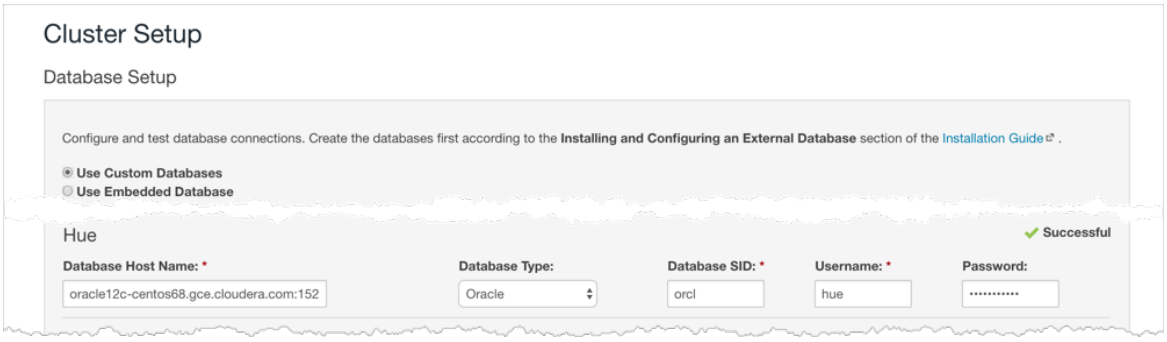


3. Stop at **Database Setup** to set connection properties (**Cluster Setup**, step 3).
   a. Select **Use Custom Database**.
   b. Under **Hue**, set the connection properties to the Oracle database.

> **Note:** Copy and store the password for the Hue embedded database (just in case).

```
Database Hostname (and port): <fqdn of host with Oracle server>:1521
Database Type (or engine): Oracle
Database SID (or name): orcl
Database Username: hue
Database Password: <hue database password>
```

   c. Click **Test Connection** and click **Continue** when successful.



4. Continue with the installation and click **Finish** to complete.
5. Add support for a multi-threaded environment:
   a. Go to **Clusters** > **Hue** > **Configuration**.
   b. Filter by Category, **Hue-service** and Scope, **Advanced**.
   c. Add support for a multi-threaded environment by setting **Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini**:

```
[desktop]
[[database]]
options={"threaded":true}
```

   d. Click **Save Changes**.

6. Restart the Hue service: select **Actions** > **Restart** and click **Restart**.
7. Log on to Hue by clicking **Hue Web UI**.

Existing CDH Installation

If you are not migrating the current (or old) database, simply connect to your new Oracle database and restart Hue. See Connect Hue to an External Database.

1. [migration only] **Stop Hue Service**
   a. In Cloudera Manager, navigate to **Cluster** > **Hue**.
   b. Select **Actions** > **Stop**.

> **Note:** If necessary, refresh the page to ensure the Hue service is stopped: .

2. [migration only] **Dump Current Database**
   a. Select **Actions** > **Dump Database**.
   b. Click **Dump Database**. The file is written to `/tmp/hue_database_dump.json` on the host of the Hue server.
   c. Log on to the *host of the Hue server* in a command-line terminal.
   d. Edit `/tmp/hue_database_dump.json` by removing all objects with `useradmin.userprofile` in the `model` field. For example:

```
# Count number of objects
grep -c useradmin.userprofile /tmp/hue_database_dump.json
```

```
vi /tmp/hue_database_dump.json
```

```
{
  "pk": 1,
  "model": "useradmin.userprofile",
  "fields": {
    "last_activity": "2016-10-03T10:06:13",
    "creation_method": "HUE",
    "first_login": false,
    "user": 1,
    "home_directory": "/user/admin"
  }
},
{
  "pk": 2,
  "model": "useradmin.userprofile",
  "fields": {
    "last_activity": "2016-10-03T10:27:10",
    "creation_method": "HUE",
    "first_login": false,
    "user": 2,
    "home_directory": "/user/alice"
  }
},
```

3. **Connect to New Database**
   a. Configure Database connections: Go to **Hue** > **Configuration**, filter by **Database**, set properties, and click **Save Changes**:

```
Hue Database Type (or engine): Oracle
Hue Database Hostname: <fqdn of host with Oracle server>
Hue Database Port: 1521
Hue Database Username: hue
Hue Database Password: <hue database password>
Hue Database Name (or SID): orcl
```

b. Add support for a multi-threaded environment: Filter by **Hue-service**, set **Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini**, and click **Save Changes**:

```
[desktop]
[[database]]
options={"threaded":true}
```

4. [migration only] **Synchronize New Database**
   a. Select **Actions** > **Synchronize Database**
   b. Click **Synchronize Database**.
5. [migration only] **Load Data from Old Database**

Important: All user tables in the Hue database must be empty. You cleaned them at step 3 of Create Hue Database. Ensure they are still clean.

```
sqlplus hue/<hue_password> < delete_from_tables.ddl
```

6. **Re/Start Hue service**
   a. Navigate to **Cluster** > **Hue**.
   b. Select **Actions** > **Start**, and click **Start**.
   c. Click **Hue Web UI** to log on to Hue with a custom Oracle database.

Configuring an Oracle Database for Cloudera Manager

After completing the above instructions to install and configure Oracle databases for Cloudera software, continue to Step 5: Set up the Cloudera Manager Database on page 115 to configure a database for Cloudera Manager.

Configuring an External Database for Sqoop 2

Sqoop 2 has a built-in Derby database, but Cloudera recommends that you use a PostgreSQL database instead, for the following reasons:

- Derby runs in embedded mode and it is not possible to monitor its health.
- Though it might be possible, Cloudera currently has no live backup strategy for the embedded Derby database.
- Under load, Cloudera has observed locks and rollbacks with the embedded Derby database that do not happen with server-based databases.

See CDH and Cloudera Manager Supported Databases for tested database versions.

Note:

Cloudera currently has no recommended way to migrate data from an existing Derby database into the new PostgreSQL database.

Use the procedure that follows to configure Sqoop 2 to use PostgreSQL instead of Apache Derby.

Install PostgreSQL

See Install and Configure PostgreSQL for Cloudera Software on page 100.

Create the Sqoop 2 User and Sqoop 2 Database

```
$ psql -U postgres
Password for user postgres: *****
```

```
postgres=# CREATE ROLE sqoop LOGIN ENCRYPTED PASSWORD 'sqoop'
 NOSUPERUSER INHERIT CREATEDB NOCREATEROLE;
CREATE ROLE

postgres=# CREATE DATABASE "sqoop" WITH OWNER = sqoop
 ENCODING = 'UTF8'
 TABLESPACE = pg_default
 LC_COLLATE = 'en_US.UTF8'
 LC_CTYPE = 'en_US.UTF8'
 CONNECTION LIMIT = -1;
CREATE DATABASE

postgres=# \q
```

Configure Sqoop 2 to use PostgreSQL

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

1. Go to the Sqoop 2 service.
2. Click the **Configuration** tab.
3. Select **Scope** > **Sqoop 2 Server**.
4. Select **Category** > **Database**.
5. Set the following properties:
   - Sqoop Repository Database Type - postgresql
   - Sqoop Repository Database Host - the hostname on which you installed the PostgreSQL server. If the port is non-default for your database type, use host:port notation.
   - Sqoop Repository Database Name, User, Password - the properties you specified in Create the Sqoop 2 User and Sqoop 2 Database on page 114.
6. Click **Save Changes** to commit the changes.
7. Restart the service.

Step 5: Set up the Cloudera Manager Database

Cloudera Manager Server includes a script that can create and configure a database for itself. The script can:

- Create the Cloudera Manager Server database configuration file.
- (MariaDB, MySQL, and PostgreSQL) Create and configure a database for Cloudera Manager Server to use.
- (MariaDB, MySQL, and PostgreSQL) Create and configure a user account for Cloudera Manager Server.

Although the script can create a database, the following procedures assume that you have already created the database as described in Step 4: Install and Configure Databases on page 87.

The following sections describe the syntax for the script and demonstrate how to use it:

Syntax for `scm_prepare_database.sh`

The syntax for the `scm_prepare_database.sh` script is as follows:

```
/usr/share/cmf/schema/scm_prepare_database.sh [options] <databaseType> <databaseName>
<databaseUser> <password>
```

Note: You can also run `scm_prepare_database.sh` without options to see the syntax.

To create a new database, you must specify the `-u` and `-p` parameters for a user with privileges to create databases. If you have already created the database as instructed in Step 4: Install and Configure Databases on page 87, do not specify these options.

The following tables describe the parameters and options for the `scm_prepare_database.sh` script:

Table 18: Parameters

Parameter (Required in bold)	Description
<databaseType>	One of the supported database types: • MariaDB: `mysql` • MySQL: `mysql` • Oracle: `oracle` • PostgreSQL: `postgresql`
<databaseName>	The name of the Cloudera Manager Server database to use. For MySQL, MariaDB, and PostgreSQL databases, the script can create the specified database if you specify the `-u` and `-p` options with the credentials of a user that has privileges to create databases and grant privileges. The default database name provided in the Cloudera Manager configuration settings is `scm`, but you are not required to use it.
<databaseUser>	The username for the Cloudera Manager Server database to create or use. The default username provided in the Cloudera Manager configuration settings is `scm`, but you are not required to use it.
<password>	The password for the *<databaseUser>* to create or use. If you do not want the password visible on the screen or stored in the command history, do not specify the password, and you are prompted to enter it as follows: `Enter SCM password:`

Table 19: Options

Option	Description
`-?|--help`	Display help.
`--config-path`	The path to the Cloudera Manager Server configuration files. The default is `/etc/cloudera-scm-server`.
`-f|--force`	If specified, the script does not stop if an error occurs.
`-h|--host`	The IP address or hostname of the host where the database is installed. The default is to use `localhost`.
`-p|--password`	The admin password for the database application. Use with the `-u` option. The default is no password. Do not put a space between `-p` and the password (for example, `-phunter2`). If you do not want the password visible on the screen or stored in the command history, use the `-p` option without specifying a password, and you are prompted to enter it as follows: `Enter database password:` If you have already created the database, do not use this option.
`-P|--port`	The port number to use to connect to the database. The default port is 3306 for MariaDB, 3306 for MySQL, 5432 for PostgreSQL, and 1521 for Oracle. This option is used for a remote connection only.

Option	Description
`--scm-host`	The hostname where the Cloudera Manager Server is installed. If the Cloudera Manager Server and the database are installed on the same host, do not use this option or the `-h` option.
`--scm-password-script`	A script to execute whose `stdout` provides the password for user SCM (for the database).
`-u|--user`	The admin username for the database application. Use with the `-p` option. Do not put a space between `-u` and the username (for example, `-uroot`). If this option is supplied, the script creates a user and database for the Cloudera Manager Server. If you have already created the database, do not use this option.

Preparing the Cloudera Manager Server Database

1. Run the `scm_prepare_database.sh` script on the Cloudera Manager Server host, using the database name, username, and password you created in Step 4: Install and Configure Databases on page 87:

```
sudo /usr/share/cmf/schema/scm_prepare_database.sh <databaseType> <databaseName>
<databaseUser>
```

   When prompted, enter the password.

2. If it exists, remove the embedded PostgreSQL properties file:

```
sudo rm /etc/cloudera-scm-server/db.mgmt.properties
```

The following examples demonstrate the syntax and output of the `scm_prepare_database.sh` script for different scenarios:

Example 1: Running the script when MySQL or MariaDB is co-located with the Cloudera Manager Server

This example assumes that you have already created the Cloudera Management Server database and database user, naming both `scm`:

```
sudo /usr/share/cmf/schema/scm_prepare_database.sh mysql scm scm
```

```
Enter SCM password:
JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing:  /usr/java/jdk1.7.0_67-cloudera/bin/java -cp
/usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-java.jar:/usr/share/java/postgresql-connector-java.jar:/usr/share/cmf/schema/../lib/*
 com.cloudera.enterprise.dbutil.DbCommandExecutor /etc/cloudera-scm-server/db.properties
 com.cloudera.cmf.db.
[                          main] DbCommandExecutor              INFO  Successfully
connected to database.
All done, your SCM database is configured correctly!
```

Example 2: Running the script when MySQL or MariaDB is installed on another host

This example demonstrates how to run the script on the Cloudera Manager Server host (`cm01.example.com`) and connect to a remote MySQL or MariaDB host (`db01.example.com`):

```
sudo /usr/share/cmf/schema/scm_prepare_database.sh mysql -h db01.example.com --scm-host
 cm01.example.com scm scm
```

```
Enter database password:
JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing:  /usr/java/jdk1.7.0_67-cloudera/bin/java -cp
/usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-java.jar:/usr/share/java/postgresql-connector-java.jar:/usr/share/cmf/schema/../lib/*
```

```
 com.cloudera.enterprise.dbutil.DbCommandExecutor /etc/cloudera-scm-server/db.properties
 com.cloudera.cmf.db.
[                          main] DbCommandExecutor              INFO  Successfully
connected to database.
All done, your SCM database is configured correctly!
```

Example 3: Running the script to configure Oracle

```
sudo /usr/share/cmf/schema/scm_prepare_database.sh -h cm-oracle.example.com oracle orcl
 sample_user sample_pass
```

```
Enter SCM password:
JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing:  /usr/java/jdk1.7.0_67-cloudera/bin/java -cp
/usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-java.jar:/usr/share/java/postgresql-connector-java.jar:/usr/share/cmf/schema/../lib/*
 com.cloudera.enterprise.dbutil.DbCommandExecutor /etc/cloudera-scm-server/db.properties
 com.cloudera.cmf.db.
[                          main] DbCommandExecutor              INFO  Successfully
connected to database.
All done, your SCM database is configured correctly!
```

Installing CDH

After configuring the Cloudera Manager Server database, continue to Step 6: Install CDH and Other Software on page 118.

Step 6: Install CDH and Other Software

After setting up the Cloudera Manager database, start Cloudera Manager Server, and log in to the Cloudera Manager Admin Console:

1. Start Cloudera Manager Server:
   - **RHEL 7 compatible:**

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

   - **RHEL 6 compatible, Ubuntu, SLES:**

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

2. Wait several minutes for the Cloudera Manager Server to start. To observe the startup process, run the following on the Cloudera Manager Server host:

```
sudo tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log
```

   When you see this log entry, the Cloudera Manager Admin Console is ready:

```
INFO WebServerImpl:com.cloudera.server.cmf.WebServerImpl: Started Jetty server.
```

   If the Cloudera Manager Server does not start, see Troubleshooting Installation Problems on page 152.

3. In a web browser, go to `http://<server_host>:7180`, where *<server_host>* is the FQDN or IP address of the host where the Cloudera Manager Server is running.
4. Log into Cloudera Manager Admin Console. The default credentials are:

   Username: `admin`

   Password: `admin`

> **Note:** Cloudera Manager does not support changing the `admin` username for the installed account. You can change the password using Cloudera Manager after you run the installation wizard. Although you cannot change the `admin` username, you can add a new user, assign administrative privileges to the new user, and then delete the default `admin` account.

5. After you log in, the **End User License Terms and Conditions** page displays. Read the terms and conditions and then check the box labeled **Yes, I accept the End User License Terms and Conditions** to accept them. Click **Continue**, and the installation wizard launches.

The following sections guide you through each step of the installation wizard:

Welcome to Cloudera Manager

From the **Welcome to Cloudera Manager** page, you can select the edition of Cloudera Manager to install and, optionally, install a license:

1. Choose which edition to install:
   - Cloudera Express, which does not require a license, but provides a limited set of features.
   - Cloudera Enterprise Cloudera Enterprise Trial, which does not require a license, but expires after 60 days and cannot be renewed.
   - Cloudera Enterprise with one of the following license types:
     - Basic Edition
     - Flex Edition
     - Cloudera Enterprise

   If you choose Cloudera Express or Cloudera Enterprise Cloudera Enterprise Trial, you can upgrade the license at a later time. For more information, see Managing Licenses.
2. If you select Cloudera Enterprise, install a license:
   a. Click the **Select License File** text field.
   b. Browse to the location of your license file, select the file, and then click **Open**.
   c. Click the **Upload** button.
3. Information is displayed indicating what the CDH installation includes. At this point, you can click the **Support** drop-down menu to access online Help or the Support Portal.
4. Click **Continue** to proceed with the installation.

Thank you for choosing Cloudera Manager and CDH

The **Thank you for choosing Cloudera Manager and CDH** page lists the software that is available to be installed. Click **Continue** to proceed with the installation.

Specify hosts for your CDH cluster installation

Choose which hosts will run CDH and managed services

1. Do one of the following depending on whether you are using Cloudera Manager to install software:
   - If you are using Cloudera Manager to install software, search for and choose hosts:
     1. To enable Cloudera Manager to automatically discover hosts on which to install CDH and managed services, enter the cluster hostnames or IP addresses. You can also specify hostname and IP address ranges. For example:

Range Definition	Matching Hosts
10.1.1.[1-4]	10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4
host[1-3].example.com	host1.example.com, host2.example.com, host3.example.com

Range Definition	Matching Hosts
host[07-10].example.com	host07.example.com, host08.example.com, host09.example.com, host10.example.com

Note: Unqualified hostnames (short names) must be unique in a Cloudera Manager instance. For example, you cannot have both *host01.example.com* and *host01.standby.example.com* managed by the same Cloudera Manager Server.

You can specify multiple addresses and address ranges by separating them with commas, semicolons, tabs, or blank spaces, or by placing them on separate lines. Use this technique to make more specific searches instead of searching overly wide ranges. The scan results will include all addresses scanned, but only scans that reach hosts running SSH will be selected for inclusion in your cluster by default. If you do not know the IP addresses of all of the hosts, you can enter an address range that spans over unused addresses and then clear the hosts that do not exist (and are not discovered) later in this procedure. However, keep in mind that wider ranges will require more time to scan.

2. Click **Search**. Cloudera Manager identifies the hosts on your cluster to allow you to configure them for services. If there are a large number of hosts on your cluster, wait a few moments to allow them to be discovered and shown in the wizard. If the search is taking too long, you can stop the scan by clicking **Abort Scan**. To find additional hosts, click **New Search**, add the host names or IP addresses and click **Search** again. Cloudera Manager scans hosts by checking for network connectivity. If there are some hosts where you want to install services that are not shown in the list, make sure you have network connectivity between the Cloudera Manager Server host and those hosts. Common causes of loss of connectivity are firewalls and interference from SELinux.
3. Verify that the number of hosts shown matches the number of hosts where you want to install services. Clear host entries that do not exist and clear the hosts where you do not want to install services.

2. Click **Continue**.

The **Select Repository** screen displays.

Select Repository

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

1. Select the repository type to use for the installation. In the **Choose Method** section select one of the following:
   - **Use Parcels (Recommended)**

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

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

2. Select the version of CDH to install.

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

a. If you selected **Use Parcels (Recommended)** and you do not see the version you want to install, click the **More Options** button to add the repository URL for your version. Repository URLs for CDH 5 are documented in CDH Download Information. After adding the repository, click **Save Changes** and wait a few seconds for the version to appear. If your Cloudera Manager host uses an HTTP proxy, click the **Proxy Settings** button to configure your proxy.
b. If you selected **Use Packages**, and the version you want to install is not listed, you can select **Custom Repository** to specify a repository that contains the desired version. Repository URLs for CDH 5 version are documented in CDH Download Information.

3. If you selected **Use Parcels (Recommended)**, specify any **Additional Parcels** you want to install.
4. Select the specific release of the Cloudera Manager Agent you want to install on your hosts. You can choose either the version that matches the Cloudera Manager Server you are currently using (recommended) or specify a different version in a custom repository. If you opted to use custom repositories for installation files, you can provide a GPG key URL that applies for all repositories. Repository and GPG key URLs are documented in CDH Download Information.
5. Click **Continue**.

The **Accept JDK License** screen displays.

Accept JDK License

Note: Cloudera, Inc. acquired Oracle JDK software under the Oracle Binary Code License Agreement. Pursuant to Item D(v)(a) of the SUPPLEMENTAL LICENSE TERMS of the Oracle Binary Code License Agreement, use of JDK software is governed by the terms of the Oracle Binary Code License Agreement. By installing the JDK software, you agree to be bound by these terms. If you do not wish to be bound by these terms, then do not install the Oracle JDK.

To allow Cloudera Manager to automatically install the Oracle JDK on cluster hosts, read the JDK license and check the box labeled **Install Oracle Java SE Development Kit (JDK 7)** if you accept the terms. If you installed your own Oracle JDK version in Step 2: Install Java Development Kit on page 84, leave the box unchecked.

If you allow Cloudera Manager to install the JDK, a second checkbox appears, labeled **Install Java Unlimited Strength Encryption Policy Files**. These policy files are required to enable AES-256 encryption in JDK versions lower than 1.8u161. JDK 1.8u161 and higher enable unlimited strength encryption by default, and do not require policy files.

After reading the license terms and checking the applicable boxes, click **Continue**.

The **Single User Mode** page displays.

Single User Mode

The **Single User Mode** page allows you to enable single user mode. Single user mode is not recommended for most environments. If you need to enable single user mode, check the box labeled **Enable Single User Mode**. Otherwise, leave it unchecked, and click **Continue**. The **Enter Login Credentials** page displays.

Enter Login Credentials

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

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

The **Install Agents** page displays.

Install Agents

The **Install Agents** page displays the progress of the installation. You can click on the **Details** link for any host to view the installation log. If the installation is stalled, you can click the **Abort Installation** button to cancel the installation and then view the installation logs to troubleshoot the problem.

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

If you selected the option to manually install agents, see Manually Install Cloudera Software Packages on page 50 for the procedure and then continue with the next steps on this page.

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

If you are using parcels, the **Install Parcels** page displays. If you chose to install using packages, the **Inspect Hosts** page displays.

Install Parcels

If you selected parcels for the installation method, the **Install Parcels** page reports the installation progress of the parcels you selected earlier. After the parcels are downloaded, progress bars appear representing each cluster host. You can click on an individual progress bar for details about that host.

After the installation is complete, click **Continue**.

The **Inspect hosts for correctness** page displays.

Inspect hosts for correctness

The **Inspect Hosts** page runs the Host Inspector to search for common configuration problems. View the results and address any identified problems. Click the **Run Again** button to update the results after making any changes.

After addressing identified problems, click **Finish**.

This completes the **Cluster Installation** wizard and launches the **Cluster Setup** wizard.

Continue to Step 7: Set Up a Cluster Using the Wizard on page 122.

Step 7: Set Up a Cluster Using the Wizard

After completing the **Cluster Installation** wizard, the **Cluster Setup** wizard automatically starts. The following sections guide you through each page of the wizard:

Select Services

The **Select Services** page allows you to select the services you want to install and configure. Make sure that you have the appropriate license key for the services you want to use. You can choose from:

- **Core Hadoop**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, and Hue

- **Core with HBase**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and HBase

- **Core with Impala**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Impala

- **Core with Search**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Solr

- **Core with Spark**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Spark

- **All Services**

  HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, HBase, Impala, Solr, Spark, and Key-Value Store Indexer

- **Custom Services**

  Choose your own services. Services required by chosen services will automatically be included. Flume can be added after your initial cluster has been set up.

To include Cloudera Navigator data management, check the box labeled **Include Cloudera Navigator**.

After selecting the services you want to add, click **Continue**. The **Assign Roles** page displays.

Assign Roles

The **Assign Roles** page suggests role assignments for the hosts in your cluster. You can click on the hostname for a role to select a different host. You can also click the **View By Host** button to see all the roles assigned to a host.

To review the recommended role assignments, see Recommended Cluster Hosts and Role Distribution on page 39.

After assigning all of the roles for your services, click **Continue**. The **Setup Database** page displays.

Setup Database

On the **Setup Database** page, you can enter the database names, usernames, and passwords you created in Step 4: Install and Configure Databases on page 87.

Select the database type and enter the database name, username, and password for each service. Click **Test Connection** to validate the settings. If the connection is successful, a green checkmark and the word **Successful** appears next to each service. If there are any problems, the error is reported next to the service that failed to connect.

After verifying that each connection is successful, click **Continue**. The **Review Changes** page displays.

Review Changes

The **Review Changes** page lists default and suggested settings for several configuration parameters, including data directories.

Warning: Do not place DataNode data directories on NAS devices. When resizing an NAS, block replicas can be deleted, which results in missing blocks.

Review and make any necessary changes, and then click **Continue**. The **First Run Command** page displays.

First Run Command

The **First Run Command** page lists the details of the **First Run** command. You can expand the running commands to view the details of any step, including log files and command output. You can filter the view by selecting **Show All Steps**, **Show Only Failed Steps**, or **Show Running Steps**.

After the **First Run** command completes, click **Continue**.

Congratulations!

The **Congratulations!** page reports the success of the setup wizard. Click **Finish** to complete the wizard. The installation is complete.

Cloudera recommends that you change the default password as soon as possible by clicking the logged-in username at the top right of the home screen and clicking **Change Password**.

Installing the Cloudera Navigator Data Management Component

The Cloudera Navigator data management component is implemented in two distinct roles—Navigator Audit Server, and Navigator Metadata Server—that run on the Cloudera Management Service. These roles can be added during the initial Cloudera Manager installation, or added later to an existing Cloudera Manager cluster.

> **Important:** Cloudera Navigator Data Management requires a Cloudera Enterprise license. This feature is not available in Cloudera Express. See Managing Licenses for details.

The steps on this page are for installing Cloudera Navigator as part of a new Cloudera Manager cluster installation and for adding the service to an existing cluster. For information about upgrading an existing deployment, see Upgrading the Cloudera Navigator Data Management Component.

> **Note:** See Product Compatibility Matrix for Cloudera Navigator for information on compatible Cloudera Navigator and Cloudera Manager versions.

Minimum Recommended Memory and Disk Space

Resource	Navigator Audit Server	Navigator Metadata Server
Memory	Varies, but requires less than Navigator Metadata Server	40 GB total
Java heap size	2 – 3 GB	10–20 GB (initial setup)
OS buffer cache	20 GB	20 GB (initial setup). Increase by 20-GB increments over time as needed.
Disk	Multiple hundreds of GB. Depends on cluster size and audit volumes generated.	200 GB (SSD recommended)
Default path	None. Location of the Cloudera Navigator database.	`/var/lib/cloudera-scm-navigator`

Navigator Metadata Server and Navigator Audit Server have different recommended configurations that you should consider when you plan your deployment. For initial installation, keep the following in mind:

- **Navigator Audit Server Memory and Disk Requirements**—For Navigator Audit Server, a Java heap size of 2-3 GB (gigabytes) is usually sufficient (memory typically does not pose any issues). For Navigator Audit Server, it is the database configuration that can affect performance and so must be configured properly. Because Navigator Audit Server might need to push millions of rows of audit data daily (depending on the cluster size, number of services, and other factors), Cloudera recommends:
  - Set up the database on the same host as the Navigator Audit Server to minimize latency.
  - Monitor the database workload over time and tune as needed.
- **Navigator Metadata Server Memory and Disk Requirements**—Navigator Metadata Server relies on an embedded Solr instance for its Search capability. The Solr indexes are saved locally to the host's hard-disk drive and typically consume only tens of GBs of disk space, so allocating ~200 GBs for the data is usually sufficient. For Navigator Metadata Server disk, Cloudera recommends:
  - Mount SSD drives on the host where the Solr index will be located, for fastest I/O.
  - Use the Purge function once the system is up and running to keep the hard-disk drive consumption at that location in check.

Bottlenecks that might emerge for Navigator Metadata Server are typically associated with I/O and memory (not CPU). Memory includes Java heap size and available RAM that can be used for the OS buffer cache setting. For Navigator Metadata Server RAM, Cloudera recommends:

- Set Java heap size to 10-20 GB, which should be sufficient for initial setup.
- Increase the OS buffer cache by 20 GB to improve performance if necessary, depending on the cluster activity.

See Navigator Metadata Server Tuning for more information.

Configuring a Database for Cloudera Navigator

During the Cloudera Navigator installation process, you must select a database to store audit events and policy, role, and audit report metadata. You can choose the embedded PostgreSQL database, or you can choose an external database such as Oracle or MySQL (see CDH and Cloudera Manager Supported Databases for other supported database systems).

For production environments, Cloudera recommends using an external database rather than the embedded PostgreSQL database. In addition, the database must be setup and running before you begin the installation process. For more information, see Step 4: Install and Configure Databases on page 87.

Adding Cloudera Navigator Roles During the Cloudera Manager Installation Process

Cloudera Manager Required Role: Full Administrator

1. Install Cloudera Manager as detailed in Cloudera Installation Guide on page 8.
2. On the first page of the Cloudera Manager installation wizard, choose one of the license options that supports Cloudera Navigator:
   - Cloudera Enterprise Cloudera Enterprise Trial
   - Cloudera Enterprise, and upload the license:
     1. Click **Upload License**.
     2. Click the document icon to the left of the **Select a License File** text field.
     3. Go to the location of your license file, click the file, and click **Open**.
     4. Click **Upload**.
   - Flex Edition
   - Cloudera Enterprise
3. Click **Continue** to proceed with the installation.
4. In the first page of the **Add Services** procedure, click the **Include Cloudera Navigator** checkbox.
5. To use external databases, enter the Cloudera Navigator Audit Server and Metadata Server database properties in the **Database Setup** page.

Adding Cloudera Navigator Data Management Roles to an Existing Cloudera Manager Cluster

If the Cloudera Manager cluster has sufficient resources, you can add instances of either Cloudera Navigator roles to the cluster at any time. For more information, see:

- Adding the Navigator Audit Server Role
- Adding the Navigator Metadata Server Role

Cloudera Navigator Data Management Documentation

Other topics related to configuring, upgrading, managing, and using Cloudera Navigator Data Management component are listed in the following table.

FAQ	Cloudera Navigator Frequently Asked Questions answers common questions about Cloudera Navigator data management component and how it interacts with other Cloudera products and cluster components.

Introduction	Cloudera Navigator Data Management Overview provides an overview for data stewards, governance and compliance teams, data engineers, and administrators. Includes Getting Started with Cloudera Navigator, an overview of the Cloudera Navigator console (the UI) and the Cloudera Navigator APIs.
User Guide	Cloudera Navigator Data Management guide shows data stewards, compliance officers, and other business users how to use Cloudera Navigator for data governance, compliance, data stewardship, and other tasks. Topics include Auditing, Metadata, Lineage Diagrams, Cloudera Navigator and the Cloud, Services and Security Management, and more.
Upgrade	Upgrading the Cloudera Navigator Data Management Component
Security	Configuring Authentication for Cloudera Navigator
	Configuring TLS/SSL for Navigator Audit Server
	Configuring TLS/SSL for Navigator Metadata Server
Release Notes	Cloudera Navigator Data Management Release Notes

Installing Cloudera Navigator Encryption Components

The following sections demonstrate how to install the Cloudera Navigator encryption components, used for encrypting data at rest in Cloudera Enterprise:

Installing Cloudera Navigator Key Trustee Server

Important: Before installing Cloudera Navigator Key Trustee Server, see Encrypting Data at Rest for important considerations.

When the Key Trustee Server role is created it is tightly bound to the identity of the host on which it is installed. Moving the role to a different host, changing the host name, or changing the IP of the host is *not* supported

You can install Navigator Key Trustee Server using Cloudera Manager with parcels or using the command line with packages. See Parcels for more information on parcels.

Note: If you are using or planning to use Key Trustee Server in conjunction with a CDH cluster, Cloudera strongly recommends using Cloudera Manager to install and manage Key Trustee Server to take advantage of Cloudera Manager's robust deployment, management, and monitoring capabilities.

Prerequisites

See Data at Rest Encryption Requirements for more information about encryption and Key Trustee Server requirements.

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:

- Using an Internal Parcel Repository on page 45
- Using an Internal Package Repository on page 48

Installing Key Trustee Server

Important: This feature requires a Cloudera Enterprise license. It is not available in Cloudera Express. See Managing Licenses for more information.

Installing Key Trustee Server Using Cloudera Manager

Note: These instructions apply to using Cloudera Manager only. To install Key Trustee Server using packages, skip to Installing Key Trustee Server Using the Command Line on page 128.

If you are installing Key Trustee Server for use with HDFS Transparent Encryption, the **Set up HDFS Data At Rest Encryption** wizard installs and configures Key Trustee Server. See Enabling HDFS Encryption Using the Wizard for instructions.

1. **(Recommended)** Create a new cluster in Cloudera Manager containing only the host that Key Trustee Server will be installed on. Cloudera recommends that each cluster use its own KTS instance. Although sharing a single KTS across clusters is technically possible, it is *neither approved nor supported* for security reasons—specifically, the

increased security risks associated with single point of failure for encryption keys used by multiple clusters. For a better understanding of additional security reasons for this recommendation, see Data at Rest Encryption Reference Architecture. See Adding and Deleting Clusters for instructions on how to create a new cluster in Cloudera Manager.

> **Important:** The **Add Cluster** wizard prompts you to install CDH and other cluster services. To exit the wizard without installing CDH, select a version of CDH to install and continue. When the installation begins, click the Cloudera Manager logo in the upper left corner and confirm you want to exit the wizard. This allows you to create the dedicated cluster with the Key Trustee Server hosts without installing CDH or other services that are not required for Key Trustee Server.

2. Add the internal parcel repository you created in Setting Up an Internal Repository on page 127 to Cloudera Manager following the instructions in Configuring Cloudera Manager Server Parcel Settings.
3. Download, distribute, and activate the 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.

After installing Key Trustee Server using Cloudera Manager, continue to Securing Key Trustee Server Host on page 130.

Installing Key Trustee Server Using the Command Line

> **Note:** These instructions apply to package-based installations using the command line only. To install Key Trustee Server using Cloudera Manager, see Installing Key Trustee Server Using Cloudera Manager on page 127.
>
> If you are using or planning to use Key Trustee Server in conjunction with a CDH cluster, Cloudera strongly recommends using Cloudera Manager to install and manage Key Trustee Server to take advantage of Cloudera Manager's robust deployment, management, and monitoring capabilities.

1. **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
```

2. **(RHEL 7 Only) Enable the `extras` Repository**

   Key Trustee Server requires the `python-flask` package. For RHEL 6, this package is provided in the EPEL repository. For RHEL 7, it is provided in the RHEL `extras` repository. To enable this repository, run the following command:

```
$ sudo subscription-manager repos --enable=rhel-7-server-extras-rpms
```

3. **Install the PostgreSQL 9.3 Repository**

   Note: Cloudera Navigator Key Trustee Server currently supports only PostgreSQL version 9.3. If you have a different version of PostgreSQL installed on the Key Trustee Server host, remove it before proceeding or select a different host on which to install Key Trustee Server.

   To install the PostgreSQL 9.3 repository, run the following command:

```
$ sudo yum install
http://yum.postgresql.org/9.3/redhat/rhel-6-x86_64/pgdg-redhat93-9.3-3.noarch.rpm
```

   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.

4. **Install the Cloudera Repository**

   Add the internal repository you created. See Modifying Clients to Use the Internal Repository on page 49 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
```

5. **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 Step 1: Configure a Repository.

6. **Install NTP**

   The Network Time Protocol (NTP) service synchronizes system time. Cloudera recommends using NTP to ensure that timestamps in system logs, cryptographic signatures, and other auditable events are consistent across systems. Install and start NTP with the following commands:

```
$ sudo yum install ntp
$ sudo service ntpd start
## For RHEL/CentOS 7, use 'sudo systemctl start ntpd' instead ##
```

7. **Install Key Trustee Server**

   Run the following command to install the Key Trustee Server:

```
$ sudo yum install keytrustee-server
```

   Installing the Key Trustee Server also installs required dependencies, including PostgreSQL 9.3. After the installation completes, confirm that the PostgreSQL version is 9.3 by running the command `createuser -V`.

8. **Configure Services to Start at Boot**

Ensure that `ntpd`, `keytrustee-db`, and `keytrusteed` start automatically at boot:

```
$ sudo chkconfig ntpd on
$ sudo chkconfig keytrustee-db on
$ sudo chkconfig keytrusteed on
```

The `chkconfig` command provides no output if successful.

> **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.

After installing Key Trustee Server, continue to Securing Key Trustee Server Host on page 130.

Securing Key Trustee Server Host

Cloudera strongly recommends securing the Key Trustee Server host to protect against unauthorized access to Key Trustee Server. Red Hat provides security guides for RHEL:

- RHEL 6 Security Guide
- RHEL 7 Security Guide

Cloudera also recommends configuring the Key Trustee Server host to allow network communication only over certain ports.

You can use the following examples to create `iptables` rules for an EDH cluster. Add any other ports required by your environment, subject to your organization security policies. Note that in this example port 5432 is the database port for the Key Trustee database on legacy machines (prior to release 5.5). Port 11371 is the current port on which Key Trustee communicates, and port 11381 is the database port. Exercise caution if blocking other ports, as this can cause a disruption in service. See Ports Used by Cloudera Manager and Cloudera Navigator on page 23 for details about ports used with the Key Trustee Server.

```
# Flush iptables
iptables -F
iptables -X

# Allow unlimited traffic on loopback (localhost) connection
iptables -A INPUT -i lo -j ACCEPT
iptables -A OUTPUT -o lo -j ACCEPT

# Allow established, related connections
iptables -A INPUT -m state --state ESTABLISHED,RELATED -j ACCEPT
iptables -A OUTPUT -m state --state ESTABLISHED,RELATED -j ACCEPT

# Open all Cloudera Manager ports to allow Key Trustee Server to work properly

iptables -A INPUT -p tcp -m tcp --dport 5432 -j ACCEPT
iptables -A INPUT -p tcp -m tcp --dport 11371 -j ACCEPT
iptables -A INPUT -p tcp -m tcp --dport 11381 -j ACCEPT

# Drop all other connections
iptables -P INPUT DROP
iptables -P OUTPUT ACCEPT
iptables -P FORWARD DROP

# Save iptables rules so that they're loaded if the system is restarted
sed 's/IPTABLES_SAVE_ON_STOP="no"/IPTABLES_SAVE_ON_STOP="yes"/' -i
/etc/sysconfig/iptables-config
sed 's/IPTABLES_SAVE_ON_RESTART="no"/IPTABLES_SAVE_ON_RESTART="yes"/' -i
/etc/sysconfig/iptables-config
```

Leveraging Native Processor Instruction Sets

AES-NI

The Advanced Encryption Standard New Instructions (AES-NI) instruction set is designed to improve the speed of encryption and decryption using AES. Some newer processors come with AES-NI, which can be enabled on a per-server basis. If you are uncertain whether AES-NI is available on a device, run the following command to verify:

```
$ grep -o aes /proc/cpuinfo
```

To determine whether the AES-NI kernel module is loaded, run the following command:

```
$ sudo lsmod | grep aesni
```

If the CPU supports AES-NI but the kernel module is not loaded, see your operating system documentation for instructions on installing the `aesni-intel` module.

Intel RDRAND

The Intel RDRAND instruction set, along with its underlying Digital Random Number Generator (DRNG), is useful for generating keys for cryptographic protocols without using `haveged`.

To determine whether the CPU supports RDRAND, run the following command:

```
$ grep -o rdrand /proc/cpuinfo
```

To enable RDRAND, install `rng-tools` version 4 or higher:

1. Download the source code:

```
$ sudo wget
http://downloads.sourceforge.net/project/gkernel/rng-tools/4/rng-tools-4.tar.gz
```

2. Extract the source code:

```
tar xvfz rng-tools-4.tar.gz
```

3. Enter the `rng-tools-4` directory:

```
$ cd rng-tools-4
```

4. Run `./configure`.
5. Run `make`.
6. Run `make install`.

Start `rngd` with the following command:

```
$ sudo rngd --no-tpm=1 -o /dev/random
```

Initializing Key Trustee Server

After installing Key Trustee Server, you must initialize it before it is operational. Continue to Initializing Standalone Key Trustee Server or Cloudera Navigator Key Trustee Server High Availability for instructions.

Installing Cloudera Navigator Key HSM

Important: Before installing Cloudera Navigator Key HSM, see Encrypting Data at Rest for important considerations.

Cloudera Navigator Key HSM is a universal hardware security module (HSM) driver that translates between the target HSM platform and Cloudera Navigator Key Trustee Server.

With Navigator Key HSM, you can use a Key Trustee Server to securely store and retrieve encryption keys and other secure objects, without being limited solely to a hardware-based platform.

Prerequisites

You must install Key HSM on the same host as Key Trustee Server. See Data at Rest Encryption Requirements for more information about encryption and Key HSM requirements.

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 Using an Internal Package Repository on page 48.

Installing Navigator Key HSM

Important: If you have implemented Key Trustee Server high availability, install and configure Key HSM on each Key Trustee Server host.

1. **Set up the Key HSM Repository**

   Download the Key HSM tarball and create a local Key HSM repository with the files from the tarball. See Setting Up an Internal Repository on page 132 above for more information.

2. **Install the Key HSM repository**

   Add the local Key HSM repository you created in Step 1. See Modifying Clients to Use the Internal Repository on page 49 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
```

3. **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 Step 1: Configure a Repository.

4. **Install Navigator Key HSM**

   Install the Navigator Key HSM package using `yum`:

```
$ sudo yum install keytrustee-keyhsm
```

   Cloudera Navigator Key HSM is installed to the `/usr/share/keytrustee-server-keyhsm` directory by default.

Installing Key Trustee KMS

Important:

Following these instructions installs the required software to add the Key Trustee KMS service to your cluster; this enables you to use Cloudera Navigator Key Trustee Server as the underlying keystore for HDFS Transparent Encryption. This *does not* install Key Trustee Server. See Installing Cloudera Navigator Key Trustee Server on page 127 for instructions on installing Key Trustee Server. You must install Key Trustee Server before installing and using Key Trustee KMS.

Also, when the Key Trustee KMS role is created, it is tightly bound to the identity of the host on which it is installed. Moving the role to a different host, changing the host name, or changing the IP of the host is *not* supported.

Key Trustee KMS is a custom Key Management Server (KMS) that uses Cloudera Navigator Key Trustee Server as the underlying keystore, instead of the file-based Java KeyStore (JKS) used by the default Hadoop KMS.

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.

Important: 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. For additional details about installing JCE, refer to Step 3: If you are Using AES-256 Encryption, Install the JCE Policy File.

The **KMS (Navigator Key Trustee)** service in Cloudera Manager 5.3 is renamed to **Key Trustee KMS** in Cloudera Manager 5.4.

Setting Up an Internal Repository

You must create an internal repository to install Key Trustee KMS. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see Using an Internal Parcel Repository on page 45 if you are using parcels, or Using an Internal Package Repository on page 48 if you are using packages.

Installing Key Trustee KMS Using Parcels

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 on page 47 for more information.
3. Download, distribute, and activate the Key Trustee KMS parcel. See Managing Parcels for detailed instructions on using parcels to install or upgrade components.

Note: The KEYTRUSTEE_SERVER parcel in Cloudera Manager is *not* the Key Trustee KMS parcel; it is the Key Trustee Server parcel. The parcel name for Key Trustee KMS is KEYTRUSTEE.

Installing Key Trustee KMS Using Packages

1. After Setting Up an Internal Repository on page 133, configure the Key Trustee KMS host to use the repository. See Modifying Clients to Use the Internal Repository on page 49 for more information.
2. Because the `keytrustee-keyprovider` package depends on the `hadoop-kms` package, you must add the CDH repository. See Step 1: Configure a Repository for instructions.

3. Install 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
```

Post-Installation Configuration

For instructions on installing Key Trustee Server and configuring Key Trustee KMS to use Key Trustee Server, see the following topics:

- Installing Cloudera Navigator Key Trustee Server on page 127
- Enabling HDFS Encryption Using the Wizard

Installing Navigator HSM KMS Backed by Thales HSM

Important: Following these instructions installs the required software to add the Navigator HSM KMS backed by Thales HSM to your cluster; this enables you to use a supported Thales HSM as the underlying keystore for HDFS Transparent Encryption.

HSM KMS backed by Thales HSM is a custom Key Management Server (KMS) that uses a supported Thales HSM as the underlying keystore, instead of the file-based Java KeyStore (JKS) used by the default Hadoop KMS.

Important: HSM KMS backed by Thales HSM is supported only in Cloudera Manager deployments. You can install the software using parcels or packages, but running HSM KMS backed by Thales HSM outside of Cloudera Manager is not supported.

Client Prerequisites

Navigator HSM KMS backed by Thales HSM is supported on Thales HSMs only. The Thales HSM client must be installed first.

The following Thales nSolo, nConnect software and firmware are required:

- Server version: 3.67.11cam4
- Firmware: 2.65.2
- Security World Version: 12.30

Before performing the Thales HSM setup, run the `nfkminfo` command to verify that Thales HSM is configured correctly.

```
$ sudo /opt/nfast/bin/nfkminfo
          World generation 2
          state       0x1727 Initialised Usable Recovery !PINRecovery !ExistingClient
                      RTC NVRAM FTO !AlwaysUseStrongPrimes SEEDebug
```

If state reports `!Usable` instead of `Usable`, then configure the Thales HSM before continuing. See the Thales product documentation for details about how to configure the Thales client.

Run the following command to manually add the KMS user to the `nfast` group:

```
usermod -a -G nfast kms
```

If you do not manually add the KMS user, installation can fail.

Setting Up an Internal Repository

You must create an internal repository to install Navigator HSM KMS backed by Thales HSM. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see Using an Internal Parcel Repository on page 45 if you are using parcels, or Using an Internal Package Repository on page 48 if you are using packages.

Installing Navigator HSM KMS Backed by Thales HSM Using Parcels

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 on page 47 for more information.
3. Download, distribute, and activate the Navigator HSM KMS parcel. See Managing Parcels for detailed instructions on using parcels to install or upgrade components.

> **Note:** The KEYTRUSTEE_SERVER parcel in Cloudera Manager is *not* the Key Trustee KMS parcel; it is the Key Trustee Server parcel. The parcel name for Navigator HSM KMS backed by Thales HMS is KEYTRUSTEE.

Installing Navigator HSM KMS Backed by Thales HSM Using Packages

1. After Setting Up an Internal Repository on page 135, configure the Navigator KMS Services backed by Thales HSM host to use the repository. See Modifying Clients to Use the Internal Repository on page 49 for more information.
2. Because the `keytrustee-keyprovider` package depends on the `hadoop-kms` package, you must add the CDH repository. See Step 1: Configure a Repository for instructions.
3. Install the `keytrustee-keyprovider` package using the appropriate command for your operating system:

> **Important:** When installing via packages, be sure to install on each and every host on which you wish to run the HSM KMS service.

- **RHEL-compatible**

```
$ sudo yum install keytrustee-keyprovider
```

Post-Installation Configuration

For instructions on configuring HSM KMS, see Enabling HDFS Encryption Using the Wizard.

Installing Navigator HSM KMS Backed by Luna HSM

> **Important:** Following these instructions installs the required software to add the Navigator KMS Services backed by Luna HSM to your cluster; this enables you to use a supported Luna HSM as the underlying keystore for HDFS Transparent Encryption.

Navigator HSM KMS backed by Luna HSM is a custom Key Management Server (KMS) that uses a supported Luna HSM as the underlying keystore, instead of the file-based Java KeyStore (JKS) used by the default Hadoop KMS.

Important: Navigator HSM KMS backed by Luna HSM is supported only in Cloudera Manager deployments. You can install the software using parcels or packages, but running Navigator HSM KMS backed by Luna HSM outside of Cloudera Manager is not supported.

Client Prerequisites

Navigator HSM KMS backed by Luna HSM is supported on Luna HSMs only. The Luna HSM client must be installed first.

For details about the required Luna software and firmware, refer to Navigator HSM KMS: Recommended Hardware and Supported Distributions.

Before performing the Luna HSM KMS setup, run the `vtl verify` command (located at `/usr/safenet/lunaclient/bin/vtl`) to verify that the Luna HSM is configured correctly. See the Luna product documentation for details about how to configure the Luna HSM client.

Setting Up an Internal Repository

You must create an internal repository to install Navigator HSM KMS backed by Luna HSM. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see Using an Internal Parcel Repository on page 45 if you are using parcels, or Using an Internal Package Repository on page 48if you are using packages.

Installing Navigator HSM KMS Backed by Luna HSM Using Parcels

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 on page 47 for more information.
3. Download, distribute, and activate the Navigator HSM KMS parcel. See Managing Parcels for detailed instructions on using parcels to install or upgrade components.

Note: The KEYTRUSTEE_SERVER parcel in Cloudera Manager is *not* the Key Trustee KMS parcel; it is the Key Trustee Server parcel. The parcel name for Navigator HSM KMS backed by Luna HSM is KEYTRUSTEE.

Installing Navigator HSM KMS Backed by Luna HSM Using Packages

1. After Setting Up an Internal Repository on page 136, configure the Navigator HSM KMS backed by Luna HSM host to use the repository. See Modifying Clients to Use the Internal Repository on page 49 for more information.
2. Because the `keytrustee-keyprovider` package depends on the `hadoop-kms` package, you must add the CDH repository. See Step 1: Configure a Repository for instructions.
3. Install the `keytrustee-keyprovider` package using the appropriate command for your operating system:

Important: When installing via packages, be sure to install on each and every host on which you wish to run the HSM KMS service.

- **RHEL-compatible**

```
$ sudo yum install keytrustee-keyprovider
```

Post-Installation Configuration

For instructions on configuring HSM KMS, see Enabling HDFS Encryption Using the Wizard.

Installing Cloudera Navigator Encrypt

Important: Before installing Cloudera Navigator Encrypt, see Encrypting Data at Rest and the Table 16 for important considerations.

Prerequisites

See Data at Rest Encryption Requirements for more information about encryption and Navigator Encrypt requirements.

Setting Up an Internal Repository

You must create an internal repository to install or upgrade Navigator Encrypt. For instructions on creating internal repositories (including Cloudera Manager, CDH, and Cloudera Navigator encryption components), see Using an Internal Package Repository on page 48.

Installing Navigator Encrypt (RHEL-Compatible)

Note: For details about supported Linux Operating Systems, refer to the Table 16.

1. **Install the Cloudera Repository**

   Add the internal repository you created. See Modifying Clients to Use the Internal Repository on page 49 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. **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 Kernel Libraries**

For Navigator Encrypt to run as a kernel module, you must download and install the kernel development headers. Each kernel module is compiled specifically for the underlying kernel version. Running as a kernel module allows Navigator Encrypt to provide high performance and completely transparency to user-space applications.

To determine your current kernel version, run `uname -r`.

To install the development headers for your current kernel version, run:

```
$ sudo yum install kernel-headers-$(uname -r) kernel-devel-$(uname -r)
```

For OL with the Unbreakable Enterprise Kernel (UEK), run:

```
$ sudo yum install kernel-uek-headers-$(uname -r) kernel-uek-devel-$(uname -r)
```

Note: For UEK3, you do not need to install `kernel-uek-headers-*`

If `yum` cannot find these packages, it displays an error similar to the following:

```
Unable to locate package <packagename>.
```

In this case, do one of the following to proceed:

- Find and install the kernel headers package by using a tool like RPM Pbone .
- Upgrade your kernel to the latest version. If you upgrade the kernel, you must reboot after upgrading and select the kernel from the grub menu to make it active.

4. **(RHEL or CentOS Only) Manually Install `dkms`**

   Because of a broken dependency in all versions of RHEL or CentOS, you must manually install the `dkms` package:

```
$ sudo yum install
http://repository.it4i.cz/mirrors/repoforge/redhat/el6/en/x86_64/rpmforge/RPMS/dkms-2.1.1.2-1.el6.rf.noarch.rpm
```

Note: This link is provided as an example for RHEL 6 only. For other versions, be sure to use the correct URL.

5. **Install Navigator Encrypt**

   Install the Navigator Encrypt client using the `yum` package manager:

```
$ sudo yum install navencrypt
```

If you attempt to install Navigator Encrypt with incorrect or missing kernel headers, you see a message like the following:

```
Building navencryptfs 3.8.0 DKMS kernel module...

###################### BUILDING ERROR ######################


Creating symlink /var/lib/dkms/navencryptfs/3.8.0/source ->
                 /usr/src/navencryptfs-3.8.0

DKMS: add completed.
Error! echo
Your kernel headers for kernel 3.10.0-229.4.2.el7.x86_64 cannot be found at
/lib/modules/3.10.0-229.4.2.el7.x86_64/build or
/lib/modules/3.10.0-229.4.2.el7.x86_64/source.
```

```
################### BUILDING ERROR ###################

Failed installation of navencryptfs 3.8.0 DKMS kernel module !
```

To recover, see Navigator Encrypt Kernel Module Setup.

Installing Navigator Encrypt (SLES)

1. **Install the Cloudera Repository**

   Add the internal repository you created. See Modifying Clients to Use the Internal Repository on page 49 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. **Install NTP**

   The Network Time Protocol (NTP) service synchronizes system time. Cloudera recommends using NTP to ensure that timestamps in system logs, cryptographic signatures, and other auditable events are consistent across systems. Install and start NTP with the following commands:

   - **SLES 11**

```
$ sudo zypper install ntp
# /etc/init.d/ntp start
```

   - **SLES 12**

```
$ sudo zypper install ntp
# service ntpd start
```

3. **Install the Kernel Module Package and Navigator Encrypt Client**

   Install the kernel module package (KMP) and Navigator Encrypt client with `zypper`:

```
$ sudo zypper install cloudera-navencryptfs-kmp-<kernel_flavor>
$ sudo zypper install navencrypt
```

   Replace *<kernel_flavor>* with the kernel flavor for your system. Navigator Encrypt supports the `default`, `xen`, and `ec2` kernel flavors.

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. **(SLES 12 only) Run `systemctl daemon-reload`**

   Due to changes in SLES 12, you must run the following command after installing Navigator Encrypt:

```
$ sudo systemctl daemon-reload
```

Installing Navigator Encrypt (Debian or Ubuntu)

1. **Install the Cloudera Repository**

   Add the internal repository you created. See Modifying Clients to Use the Internal Repository on page 49 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. **Install NTP**

   The Network Time Protocol (NTP) service synchronizes system time. Cloudera recommends using NTP to ensure that timestamps in system logs, cryptographic signatures, and other auditable events are consistent across systems. Install and start NTP with the following commands:

```
$ sudo apt-get install ntp
$ sudo /etc/init.d/ntp start
```

3. **Install Kernel Headers**

   Determine your kernel version by running `uname -r`, and install the appropriate headers:

```
$ sudo apt-get install linux-headers-$(uname -r)
```

4. **Install the Navigator Encrypt Client**

   Install Navigator Encrypt:

```
$ sudo apt-get install navencrypt
```

Post Installation

To ensure that Navigator Encrypt and NTP start after a reboot, add them to the start order with `chkconfig`:

```
$ sudo chkconfig --level 235 navencrypt-mount on
$ sudo chkconfig --level 235 ntpd on
```

Setting Up TLS for Navigator Encrypt Clients

Transport Layer Security (TLS) certificates are used to secure communication with Navigator Encrypt. Cloudera strongly recommends using certificates signed by a trusted Certificate Authority (CA).

If the TLS certificate is signed by an unrecognized CA, such as an internal CA, then you must add the root certificate to the host certificate truststore of each Navigator Encrypt client. Be aware that Navigator Encrypt uses the operating system's truststore, which is distinct from the JDK truststore used by Cloudera Manager.

To set up TLS certificates on a Navigator Encrypt client:

1. If not already installed, install the CA-certificates:

```
yum install ca-certificates
```

2. Enable the dynamic CA configuration feature:

```
update-ca-trust enable
```

3. Copy the root certificate into the host certificate truststore:

```
cp /path/to/root.pem /etc/pki/ca-trust/source/anchors/
```

4. Update the host certificate truststore:

```
update-ca-trust
```

Example

```
[root@navencrypt-1 ~]# service navencrypt-mount stop
Stopping navencrypt directories
 * Umounting /dev/nvtest/test1 ...                              [  OK  ]
 * Umounting /dev/nvtest/test2 ...                              [  OK  ]
 * Unloading module ...                                         [  OK  ]

[root@navencrypt-1 ~]# update-ca-trust enable
[root@navencrypt-1 ~]# cp dd-1.lab.usa.company.com.pem /etc/pki/ca-trust/source/anchors/
[root@navencrypt-1 ~]# update-ca-trust

[root@navencrypt-1 ~]# service navencrypt-mount start
Starting navencrypt directories
 * Mounting '/dev/nvtest/test1'                                 [  OK  ]
 * Mounting '/dev/nvtest/test2'                                 [  OK  ]
```

Entropy Requirements

Many cryptographic operations, such as those used with TLS or HDFS encryption, require a sufficient level of system entropy to ensure randomness; likewise, Navigator Encrypt needs a source of random numbers to ensure good performance. Hence, you need to make sure that the hosts running Navigator Encrypt (as well as Key Trustee Server, Key Trustee KMS) and have sufficient entropy to perform cryptographic operations.

You can check the available entropy on a Linux system by running the following command:

```
$ cat /proc/sys/kernel/random/entropy_avail
```

The output displays the entropy currently available. Check the entropy several times to determine the state of the entropy pool on the system. If the entropy is consistently low (500 or less), you must increase it by installing `rng-tools` version 4 or higher, and starting the `rngd` service.

Install rng_tools Using Package Manager

If version 4 or higher of the `rng-tools` package is available from the local package manager (`yum`), then install it directly from the package manager. If the appropriate version of `rng-tools` is unavailable, see Building rng-tools From Source on page 142.

> **Note:** If you're using RHEL 6.7 and later, or recent versions of Ubuntu, Debian, and SLES, then package manager should provide version 4.x or higher. Be sure to check the version of `rng-tools` provided by your package manager before installation to determine whether or not you need to build from source instead.

Run the following commands on RHEL 6-compatible systems:

```
$ sudo yum install rng-tools
$ sudo service rngd start
$ sudo chkconfig rngd on
```

For RHEL 7, run the following commands:

```
$ sudo yum install rng-tools
$ cp /usr/lib/systemd/system/rngd.service /etc/systemd/system/
$ systemctl daemon-reload
$ systemctl start rngd
$ systemctl enable rngd
```

Building rng-tools From Source

If you are unable to install `rng-tools` using package manager, then build from source.

> **Note:** If your package manager only offers an older version (3.x or earlier), then you must build from source.

To install and start `rngd` and build from source:

1. Download the source code:

```
$ sudo wget
http://downloads.sourceforge.net/project/gkernel/rng-tools/4/rng-tools-4.tar.gz
```

2. Extract the source code:

```
tar xvfz rng-tools-4.tar.gz
```

3. Enter the `rng-tools-4` directory:

```
$ cd rng-tools-4
```

4. Run `./configure`
5. Run `make`
6. Run `make install`

After you have installed `rng-tools`, start the `rngd` daemon by running the following command as root:

```
$ sudo rngd --no-tpm=1 -o /dev/random
```

For improved performance, Cloudera recommends configuring Navigator Encrypt to read directly from `/dev/random` instead of `/dev/urandom`.

To configure Navigator Encrypt to use `/dev/random` as an entropy source, add `--use-random` to the `navencrypt-prepare` command when you are setting up Navigator Encrypt.

Uninstalling and Reinstalling Navigator Encrypt

Uninstalling Navigator Encrypt

For RHEL-compatible OSes:

```
$ sudo yum remove navencrypt
$ sudo yum remove navencrypt-kernel-module
```

These commands remove the software itself. On RHEL-compatible OSes, the `/etc/navencrypt` directory is not removed as part of the uninstallation. Remove it manually if required.

Reinstalling Navigator Encrypt

After uninstalling Navigator Encrypt, repeat the installation instructions for your distribution in Installing Cloudera Navigator Encrypt on page 137.

When Navigator Encrypt is uninstalled, the configuration files and directories located in `/etc/navencrypt` are not removed. Consequently, you do not need to use the `navencrypt register` command during reinstallation. If you no longer require the previous installation configuration information in the directory `/etc/navencrypt`, you can remove its contents.

After Installation

The following topics describe post-installation actions, such as deploying client configuration and some simple tests to validate the installation and confirm that everything is working as expected.

Deploying Clients

Client configuration files are generated automatically by Cloudera Manager based on the services you install.

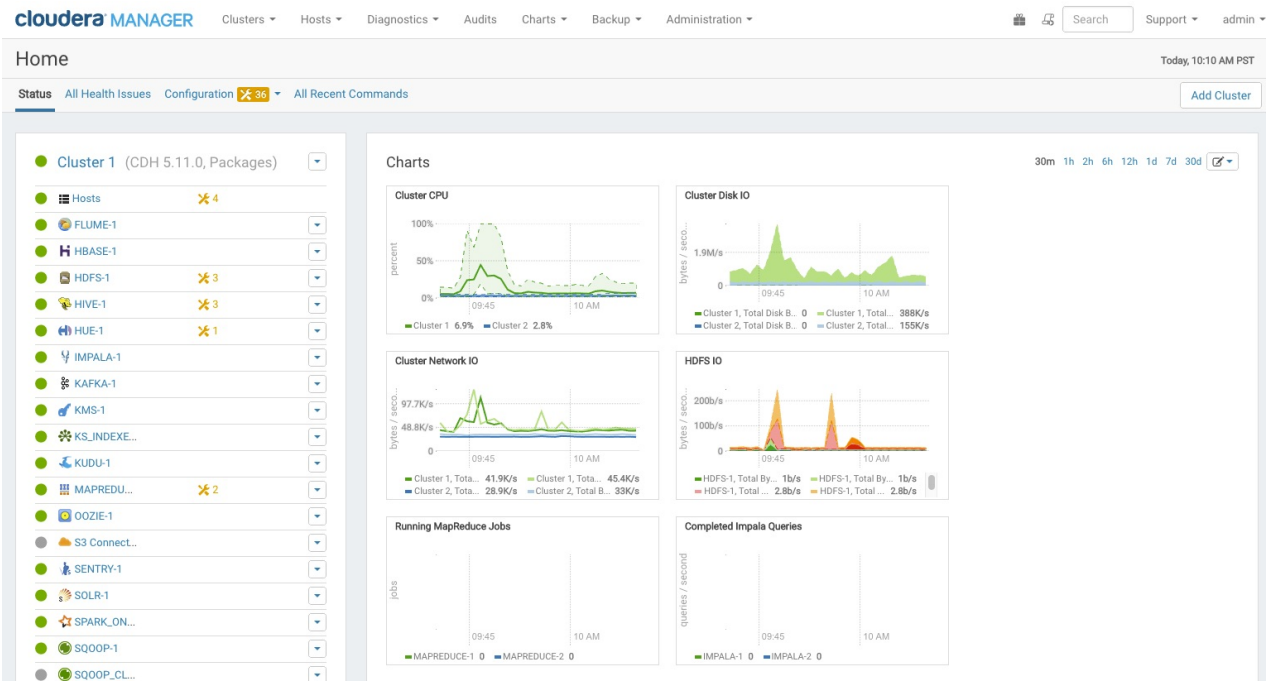
Cloudera Manager deploys these configurations automatically at the end of the installation workflow. You can also download the client configuration files to deploy them manually.

If you modify the configuration of your cluster, you might need to redeploy the client configuration files. If a service's status is "Client configuration redeployment required," you need to redeploy those files.

See Client Configuration Files for information on downloading client configuration files, or redeploying them through Cloudera Manager.

Testing the Installation

To begin testing, start the Cloudera Manager Admin Console. Once you've logged in, the Home page should look something like this:



On the left side of the screen is a list of services currently running with their status information. All the services should be running with **Good Health**. You can click each service to view more detailed information about each service. You can also test your installation by either checking each Host's heartbeats, running a MapReduce job, or interacting with the cluster with an existing Hue application.

Checking Host Heartbeats

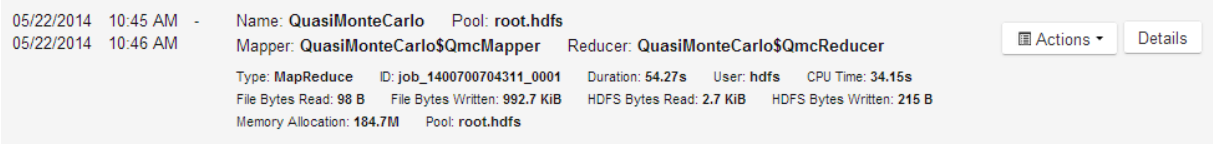
One way to check whether all the Agents are running is to look at the time since their last heartbeat. You can do this by clicking the **Hosts** tab where you can see a list of all the Hosts along with the value of their **Last Heartbeat**. By default,

every Agent must heartbeat successfully every 15 seconds. A recent value for the **Last Heartbeat** means that the Server and Agents are communicating successfully.

Running a MapReduce Job

1. Log into a host in the cluster.
2. Run the Hadoop PiEstimator example using one of the following commands:
   - **Parcel** - `sudo -u hdfs hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 100`
   - **Package** - `sudo -u hdfs hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 100`
3. Depending on whether your cluster is configured to run MapReduce jobs on the YARN or MapReduce service, view the results of running the job by selecting one of the following from the top navigation bar in the Cloudera Manager Admin Console :
   - **Clusters** > ***ClusterName*** > **yarn Applications**
   - **Clusters** > ***ClusterName*** > **mapreduce Activities**

   If you run the PiEstimator job on the YARN service (the default) you will see an entry like the following in **yarn Applications**:



Testing with Hue

A good way to test the cluster is by running a job. In addition, you can test the cluster by running one of the Hue web applications. Hue is a graphical user interface that allows you to interact with your clusters by running applications that let you browse HDFS, manage a Hive metastore, and run Hive, Impala, and Search queries, Pig scripts, and Oozie workflows.

1. In the Cloudera Manager Admin Console **Home** > **Status** tab, click the Hue service.
2. Click the **Hue Web UI** link, which opens Hue in a new window.
3. Log in with the credentials, **username:** `hdfs`, **password:** `hdfs`.
4. Choose an application in the navigation bar at the top of the browser window.

For more information, see the Hue User Guide.

Installing the GPL Extras Parcel

GPL Extras contains functionality for compressing data using the LZO compression algorithm.

To install the GPL Extras parcel:

1. Add the appropriate repository to the Cloudera Manager list of parcel repositories. Specify the repository in Cloudera Manager as follows:
   - **CDH 5:** `https://archive.cloudera.com/gplextras5/parcels/5.x.y/`

   Replace *x.y* with the minor and maintenance version (for example, `5.14.1`). If you are using LZO with Impala, make sure that you match the GPL Extras parcel version to the CDH version.
2. Download, distribute, and activate the parcel.
3. The LZO parcels require that the underlying operating system has the native LZO packages installed. If they are not installed on all cluster hosts, you can install them as follows:

RHEL compatible:

```
sudo yum install lzo
```

Debian or Ubuntu:

```
sudo apt-get install liblzo2-2
```

SLES:

```
sudo zypper install liblzo2-2
```

Migrating from Packages to Parcels

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

Managing software distribution using parcels offers many advantages over packages. To migrate from packages to the *same version* parcel, perform the following steps. To upgrade to a different version, see Upgrading CDH Using Cloudera Manager.

Download, Distribute, and Activate Parcels

1. In the Cloudera Manager Admin Console, click the parcel icon in the top navigation bar.
2. Click **Download** for the version that matches the CDH or service version of the currently installed packages. If the parcel you want is not shown here—for example, if you want to use a version of CDH that is not the most current version—you can add parcel repositories through the Parcel Configuration Settings page:
   - **CDH 5:** Impala, Kudu, Spark 1, and Search are included in the CDH parcel.
     - CDH - `https://username:password@archive.cloudera.com/p/cdh5/parcels/`
     - Accumulo--`https://username:password@archive.cloudera.com/p/accumulo-c5/parcels/`
     - GPL Extras - `https://username:password@archive.cloudera.com/p/gplextras5/parcels/`
   - **CDS 2 Powered by Apache Spark for CDH 5:**
     - The exact parcel name is dependent on the OS. You can find all the parcels at `https://username:password@archive.cloudera.com/p/spark2/parcels/`.
   - **Cloudera Navigator Key Trustee Server:**
     - Go to the Key Trustee Server download page. Select **Parcels** from the **Package or Parcel** drop-down menu, and click **DOWNLOAD NOW**. This downloads the Key Trustee Server parcels and `manifest.json` files in a `.tar.gz` file. Extract the files with the `tar xvfz filename.tar.gz` command.
   - **Cloudera Navigator Key Trustee KMS and HSM KMS:** (Cloudera Navigator HSM KMS is included in the Key Trustee KMS parcel).
     - Go to the Key Trustee KMS download page. Select **Parcels** from the **Package or Parcel** drop-down menu, and click **DOWNLOAD NOW**. This downloads the Key Trustee KMS parcels and `manifest.json` files in a `.tar.gz` file. Extract the files with the `tar xvfz filename.tar.gz` command.
   - **Sqoop Connectors:**
     - `https://archive.cloudera.com/sqoop-connectors/parcels/`

   If your Cloudera Manager Server does not have Internet access, you can obtain the required parcel file(s) and put them into a repository. See Using an Internal Parcel Repository on page 45 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**.

Restart the Cluster and Deploy Client Configuration

1. Restart the cluster:
   a. On the **Home** > **Status** tab, click

   to the right of the cluster name and select **Restart**.
   b. Click **Restart** that appears in the next screen to confirm. The **Command Details** window shows the progress of stopping services.

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

   You can optionally perform a rolling restart.
2. Redeploy client configurations:
   a. On the **Home** > **Status** tab, click

   to the right of the cluster name and select **Deploy Client Configuration**.
   b. Click **Deploy Client Configuration**.

Uninstall Packages

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:

   Warning: If you are running Key HSM, do *not* uninstall `bigtop-utils` because it is a requirement for the `keytrustee-keyhsm` package.

   - **Not including Impala, Kudu, and Search**

Operating System	Command
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, Kudu, and Search**

Operating System	Command
RHEL	`sudo yum remove 'bigtop-*' hue-common impala-shell kudu solr-server sqoop2-client hbase-solr-doc avro-libs crunch-doc avro-doc solr-doc`
SLES	`sudo zypper remove 'bigtop-*' hue-common impala-shell kudu 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 kudu 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.

Update Applications to Reference Parcel Paths

With parcels, the path to the CDH libraries is `/opt/cloudera/parcels/CDH/lib` instead of the usual `/usr/lib`. Do not link `/usr/lib/` elements to parcel-deployed paths, because the links may cause scripts that distinguish between the two paths to not work. Instead you should update your applications to reference the new library locations.

Migrating from Parcels to Packages

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

To migrate from a parcel to the *same version* packages, perform the following steps. To upgrade to a different version, see Upgrading CDH Using Cloudera Manager.

Install CDH and Managed Service Packages

Choose a Repository Strategy

To install CDH and Managed Service Packages, choose one of the following repository strategies:

- Standard Cloudera repositories. For this method, ensure you have added the required repository information to your systems.
- Internally hosted repositories. You might use internal repositories for environments where hosts do not have access to the Internet. For information about preparing your environment, see Custom Installation Solutions on page 43. When using an internal repository, you must copy the repo or list file to the Cloudera Manager Server host and update the repository properties to point to internal repository URLs.

Install CDH 5 and Managed Service Packages

Install the packages on all cluster hosts using the following steps:

- **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 kudu llama mahout oozie pig pig-udf-datafu search sentry
solr-mapreduce spark-core spark-master spark-worker 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:

- **SLES 11:**

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

- **SLES 12:**

```
$ sudo rpm --import
https://archive.cloudera.com/cdh5/sles/12/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 kudu llama mahout oozie pig pig-udf-datafu search sentry
solr-mapreduce spark-core spark-master spark-worker 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 kudu llama mahout oozie pig pig-udf-datafu search sentry
solr-mapreduce spark-core spark-master spark-worker spark-python sqoop sqoop2 whirr
```

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

Deactivate Parcels

When you deactivate a parcel, Cloudera Manager points to the installed packages, ready to be run the next time a service is restarted. To deactivate parcels,

1. Go to the Parcels page by doing one of the following:
   - Clicking the parcel indicator in the Admin Console navigation bar ( )
   - Clicking the **Hosts** in the top navigation bar, then the **Parcels** tab.
2. Click **Actions** on the activated CDH and managed service parcels and select **Deactivate**.

Restart the Cluster

1. On the **Home** > **Status** tab, click to the right of the cluster name and select **Restart**.
2. Click **Restart** that appears in the next screen to confirm. The **Command Details** window shows the progress of stopping services.

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

You can optionally perform a rolling restart.

Remove and Delete Parcels

Removing a Parcel

From the Parcels page, in the Location selector, choose ***ClusterName*** or **All Clusters**, click the to the right of an **Activate** button, and select **Remove from Hosts**.

Deleting a Parcel

From the Parcels page, in the Location selector, choose ***ClusterName*** or **All Clusters**, and click the to the right of a **Distribute** button, and select **Delete**.

Troubleshooting Installation Problems

This topic describes common installation issues and suggested solutions.

Navigator HSM KMS Backed by Thales HSM installation fails

The installation of the Navigator HSM KMS backed by Thales HSM fails with the following error message in the role log:

```
ERROR: Hadoop KMS could not be started

REASON: com.ncipher.provider.nCRuntimeException:
com.ncipher.km.nfkm.nfkmCommunicationException The nfkm command program has terminated
 unexpectedly.
```

Possible Reasons

The KMS user is not part of the `nfast` group on the host(s) running the Navigator HSM KMS backed by Thales HSM role.

Possible Solutions

Add the KMS user to the `nfast` group on the host(s) running the Navigator HSM KMS backed by Thales HSM role:

```
$ sudo usermod -G nfast kms
```

Failed to start server reported by cloudera-manager-installer.bin

"Failed to start server" reported by `cloudera-manager-installer.bin`. `/var/log/cloudera-scm-server/cloudera-scm-server.log`contains a message beginning `Caused by: java.lang.ClassNotFoundException: com.mysql.jdbc.Driver...`

Possible Reasons

You might have SELinux enabled.

Possible Solutions

Disable SELinux by running `sudo setenforce 0` on the Cloudera Manager Server host. To disable it permanently, edit `/etc/selinux/config`. For more information, see Setting SELinux mode.

Installation interrupted and installer does not restart

Installation interrupted and installer does not restart.

Possible Reasons

You need to do some manual cleanup.

Possible Solutions

See Uninstalling Cloudera Manager and Managed Software on page 156.

Cloudera Manager Server fails to start with MySQL

Cloudera Manager Server fails to start and the Server is configured to use a MySQL database to store information about service configuration.

Possible Reasons

Tables might be configured with the ISAM engine. The Server does not start if its tables are configured with the MyISAM engine, and an error such as the following appears in the log file:

```
Tables ... have unsupported engine type ... .  InnoDB is required.
```

Possible Solutions

Make sure that the InnoDB engine is configured, not the MyISAM engine. To check what engine your tables are using, run the following command from the MySQL shell: `mysql> show table status;`

For more information, see Install and Configure MySQL for Cloudera Software on page 94.

Agents fail to connect to Server

Agents fail to connect to Server. You get an Error 113 ('No route to host') in `/var/log/cloudera-scm-agent/cloudera-scm-agent.log`.

Possible Reasons

You might have SELinux or `iptables` enabled.

Possible Solutions

Check `/var/log/cloudera-scm-server/cloudera-scm-server.log` on the Server host and `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts. Disable SELinux and `iptables`. For more information, see Setting SELinux mode and Disabling the Firewall.

Cluster hosts do not appear

Some cluster hosts do not appear when you click **Find Hosts** in install or update wizard.

Possible Reasons

You might have network connectivity problems.

Possible Solutions

- Make sure all cluster hosts have SSH port 22 open.
- Check other common causes of loss of connectivity such as firewalls and interference from SELinux. For more information, see Setting SELinux mode and Disabling the Firewall.

"Access denied" in install or update wizard

"Access denied" in install or update wizard during database configuration for Activity Monitor or Reports Manager.

Possible Reasons

Hostname mapping or permissions are not set up correctly.

Possible Solutions

- For hostname configuration, see Configuring Network Names.
- For permissions, make sure the values you enter into the wizard match those you used when you configured the databases. The value you enter into the wizard as the database hostname *must* match the value you entered for the hostname (if any) when you configured the database.

  For example, if you had entered the following when you created the database

```
grant all on activity_monitor.* TO 'amon_user'@'myhost1.myco.com' IDENTIFIED BY
'amon_password';
```

  the value you enter here for the database hostname must be `myhost1.myco.com`. If you did not specify a host, or used a wildcard to allow access from any host, you can enter either the fully qualified domain name (FQDN), or `localhost`. For example, if you entered

```
grant all on activity_monitor.* TO 'amon_user'@'%' IDENTIFIED BY 'amon_password';
```

  the value you enter for the database hostname can be either the FQDN or `localhost`.

Databases fail to start.

Activity Monitor, Reports Manager, or Service Monitor databases fail to start.

Possible Reasons

MySQL `binlog` format problem.

Possible Solutions

Set `binlog_format=mixed` in `/etc/my.cnf`. For more information, see this MySQL bug report. See also Step 4: Install and Configure Databases on page 87.

Cloudera services fail to start

Cloudera services fail to start.

Possible Reasons

Java might not be installed or might be installed at a custom location.

Possible Solutions

See Configuring a Custom Java Home Location on page 66 for more information on resolving this issue.

Activity Monitor displays a status of **BAD**

The Activity Monitor displays a status of **BAD** in the Cloudera Manager Admin Console. The log file contains the following message:

```
ERROR 1436 (HY000): Thread stack overrun: 7808 bytes used of a 131072 byte stack, and
128000 bytes needed.
Use 'mysqld -O thread_stack=#' to specify a bigger stack.
```

Possible Reasons

The MySQL thread stack is too small.

Possible Solutions

1. Update the `thread_stack` value in `my.cnf` to `256KB`. The `my.cnf` file is normally located in `/etc` or `/etc/mysql`.
2. Restart the `mysql` service: `$ sudo service mysql restart`
3. Restart Activity Monitor.

Activity Monitor fails to start

The Activity Monitor fails to start. Logs contain the error `read-committed isolation not safe for the statement binlog format`.

Possible Reasons

The `binlog_format` is not set to mixed.

Possible Solutions

Modify the `mysql.cnf` file to include the entry for `binlog format` as specified in Install and Configure MySQL for Cloudera Software on page 94.

Create Hive Metastore Database Tables command fails

The **Create Hive Metastore Database Tables** command fails due to a problem with an escape string.

Possible Reasons

PostgreSQL versions 9 and higher require special configuration for Hive because of a backward-incompatible change in the default value of the `standard_conforming_strings` property. Versions up to PostgreSQL 9.0 defaulted to `off`, but starting with version 9.0 the default is on.

Possible Solutions

As the administrator user, use the following command to turn `standard_conforming_strings` off:

```
ALTER DATABASE <hive_db_name> SET standard_conforming_strings = off;
```

Oracle invalid identifier

If you are using an Oracle database and the Cloudera **Navigator Analytics** > **Audit** > **Activity** tab displays "No data available" and there is an Oracle error about "invalid identifier" with the query containing the reference to dbms_crypto in the log.

Possible Reasons

You have not granted execute permission to `sys.dbms_crypto`.

Possible Solutions

Run `GRANT EXECUTE ON sys.dbms_crypto TO` *nav*`;`, where *nav* is the user of the Navigator Audit Server database.

Uninstalling Cloudera Manager and Managed Software

Use the following instructions to uninstall the Cloudera Manager Server, Agents, managed software, and databases.

Uninstalling Cloudera Manager and Managed Software

Follow the steps in this section to remove software and data.

Record User Data Paths

The user data paths listed Remove User Data on page 159, `/var/lib/flume-ng /var/lib/hadoop* /var/lib/hue /var/lib/navigator /var/lib/oozie /var/lib/solr /var/lib/sqoop* /var/lib/zookeeper` *`data_drive_path`*`/dfs` *`data_drive_path`*`/mapred` *`data_drive_path`*`/yarn`, are the default settings. However, at some point they might have been reconfigured in Cloudera Manager. If you want to remove all user data from the cluster and have changed the paths, either when you installed CDH and managed services or at some later time, note the location of the paths by checking the configuration in each service.

Stop all Services

For each cluster managed by Cloudera Manager:

1. For each cluster managed by Cloudera Manager:
   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. On the **Home** > **Status** tab, click
   to the right of the Cloudera Management Service entry and select **Stop**. 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.

Deactivate and Remove Parcels

If you installed using packages, skip this step and go to Uninstall the Cloudera Manager Server on page 157; you will remove packages in Uninstall Cloudera Manager Agent and Managed Software on page 157. If you installed using parcels remove them as follows:

1. Click the parcel indicator in the main navigation bar.
2. In the **Location** selector on the left, select **All Clusters**.
3. For each activated parcel, select **Actions** > **Deactivate**. When this action has completed, the parcel button changes to **Activate**.
4. For each activated parcel, select **Actions** > **Remove from Hosts**. When this action has completed, the parcel button changes to **Distribute**.
5. For each activated parcel, select **Actions** > **Delete**. This removes the parcel from the local parcel repository.

There might be multiple parcels that have been downloaded and distributed, but that are not active. If this is the case, you should also remove those parcels from any hosts onto which they have been distributed, and delete the parcels from the local repository.

Delete the Cluster

On the **Home** page, Click the drop-down list next to the cluster you want to delete and select **Delete**.

Uninstall the Cloudera Manager Server

The commands for uninstalling the Cloudera Manager Server depend on the method you used to install it. Refer to steps below that correspond to the method you used to install the Cloudera Manager Server.

- **If you used the cloudera-manager-installer.bin file** - Run the following command on the Cloudera Manager Server host:

```
$ sudo /usr/share/cmf/uninstall-cloudera-manager.sh
```

- **If you did not use the cloudera-manager-installer.bin file** - If you installed the Cloudera Manager Server using a different installation method such as Puppet, run the following commands on the Cloudera Manager Server host.

  1. Stop the Cloudera Manager Server and its database:

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

  2. Uninstall the Cloudera Manager Server and its database. This process described also removes the embedded PostgreSQL database software, if you installed that option. If you did not use the embedded PostgreSQL database, omit the `cloudera-manager-server-db` steps.

     RHEL systems:

```
sudo yum remove cloudera-manager-server
sudo yum remove cloudera-manager-server-db-2
```

     SLES systems:

```
sudo zypper -n rm --force-resolution cloudera-manager-server
sudo zypper -n rm --force-resolution cloudera-manager-server-db-2
```

     Debian/Ubuntu systems:

```
sudo apt-get remove cloudera-manager-server
sudo apt-get remove cloudera-manager-server-db-2
```

Uninstall Cloudera Manager Agent and Managed Software

Do the following on all Agent hosts:

1. Stop the Cloudera Manager Agent.

   RHEL-compatible 7 and higher

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

   All other RHEL/SLES systems:

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

Debian/Ubuntu systems:

```
$ sudo /usr/sbin/service cloudera-scm-agent hard_stop
```

2. Uninstall software:

OS	Parcel Install	Package Install
RHEL	`$ sudo yum remove 'cloudera-manager-*'`	• **CDH 5** `$ sudo yum remove 'cloudera-manager-*' 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 hue-common oozie-client solr solr-doc sqoop2-client zookeeper`
SLES	`$ sudo zypper remove 'cloudera-manager-*'`	• **CDH 5** `$ sudo zypper remove 'cloudera-manager-*' 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 hue-common oozie-client solr solr-doc sqoop2-client zookeeper`
Debian/Ubuntu	`$ sudo apt-get purge 'cloudera-manager-*'`	• **CDH 5** `$ sudo apt-get purge 'cloudera-manager-*' 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 hue-common oozie-client solr solr-doc sqoop2-client zookeeper`

3. Run the `clean` command:

RHEL

```
$ sudo yum clean all
```

SLES

```
$ sudo zypper clean
```

Debian/Ubuntu

```
$ sudo apt-get clean
```

Remove Cloudera Manager and User Data

Kill Cloudera Manager and Managed Processes

On all Agent hosts, kill any running Cloudera Manager and managed processes:

```
$ for u in cloudera-scm flume hadoop hdfs hbase hive httpfs hue impala llama mapred
oozie solr spark sqoop sqoop2 yarn zookeeper; do sudo kill $(ps -u $u -o pid=); done
```

Note: This step should not be necessary if you stopped all the services and the Cloudera Manager Agent correctly.

Remove Cloudera Manager Data

If you are uninstalling on RHEL, run the following commands on all Agent hosts to permanently remove Cloudera Manager data. If you want to be able to access any of this data in the future, you must back it up before removing it. If you used an embedded PostgreSQL database, that data is stored in `/var/lib/cloudera-scm-server-db`.

```
$ sudo umount cm_processes
$ sudo rm -Rf /usr/share/cmf /var/lib/cloudera* /var/cache/yum/cloudera*
/var/log/cloudera* /var/run/cloudera*
```

Remove the Cloudera Manager Lock File

On all Agent hosts, run this command to remove the Cloudera Manager lock file:

```
$ sudo rm /tmp/.scm_prepare_node.lock
```

Remove User Data

This step permanently removes all user data. To preserve the data, copy it to another cluster using the `distcp` command before starting the uninstall process. On all Agent hosts, run the following commands:

```
$ sudo rm -Rf /var/lib/flume-ng /var/lib/hadoop* /var/lib/hue /var/lib/navigator
/var/lib/oozie /var/lib/solr /var/lib/sqoop* /var/lib/zookeeper
```

Run the following command on each data drive on all Agent hosts (adjust the paths for the data drives on each host):

```
$ sudo rm -Rf data_drive_path/dfs data_drive_path/mapred data_drive_path/yarn
```

Note: For additional information about uninstalling CDH, including clean-up of CDH files, see Uninstalling CDH Components.

Stop and Remove External Databases

If you chose to store Cloudera Manager or user data in an external database, see the database vendor documentation for details on how to remove the databases.

Uninstalling a CDH Component From a Single Host

The following procedure removes CDH software components from a single host that is managed by Cloudera Manager.

1. In the Cloudera Manager Administration Console, select the **Hosts** tab.

   A list of hosts in the cluster displays.

2. Select the host where you want to uninstall CDH software.
3. Click the **Actions for Selected** button and select **Remove From Cluster**.

   Cloudera Manager removes the roles and host from the cluster.

4. (Optional) Manually delete the `krb5.conf` file used by Cloudera Manager.

Appendix: Apache License, Version 2.0

SPDX short identifier: Apache-2.0

Apache License
Version 2.0, January 2004
http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims

licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

1. You must give any other recipients of the Work or Derivative Works a copy of this License; and
2. You must cause any modified files to carry prominent notices stating that You changed the files; and
3. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
4. If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions.

Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks.

This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty.

Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability.

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability.

While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

APPENDIX: How to apply the Apache License to your work

To apply the Apache License to your work, attach the following boilerplate notice, with the fields enclosed by brackets "[]" replaced with your own identifying information. (Don't include the brackets!) The text should be enclosed in the appropriate comment syntax for the file format. We also recommend that a file or class name and description of purpose be included on the same "printed page" as the copyright notice for easier identification within third-party archives.

```
Copyright [yyyy] [name of copyright owner]

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
```