

Installation

Date published: 2019-11-22

Date modified:



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

CDP Private Cloud Base Installation Guide.....	6
Version and Download Information.....	6
Cloudera Manager Version Information.....	6
Cloudera Manager Download Information.....	6
Cloudera Runtime Version Information.....	7
Cloudera Runtime Download Information.....	7
CDP Private Cloud Base Trial Download Information.....	7
Trial Installation.....	8
Before You Begin a Trial Installation.....	8
Installing a Trial Cluster.....	9
Step 1: Run the Cloudera Manager Server Installer.....	9
Step 2: Configure the Ranger and DAS Databases.....	10
Step 3: Install Runtime Using the Wizard.....	11
Step 4: Set Up a Cluster Using the Wizard.....	14
Stopping the Embedded PostgreSQL Database.....	16
Starting the Embedded PostgreSQL Database.....	17
Changing Embedded PostgreSQL Database Passwords.....	17
Migrating from the Cloudera Manager Embedded PostgreSQL Database Server to an External PostgreSQL Database.....	19
Prerequisites.....	19
Identify Roles that Use the Embedded Database Server.....	19
Migrate Databases from the Embedded Database Server to the External PostgreSQL Database Server.....	22
Production Installation: Before You Install.....	25
Storage Space Planning for Cloudera Manager.....	26
Cluster Lifecycle Management with Cloudera Manager.....	33
Configure Network Names.....	36
Setting SELinux Mode.....	37
Disabling the Firewall.....	38
Enable an NTP Service.....	38
Impala Requirements.....	39
sudo Commands Run by Cloudera Manager.....	41
Ports.....	41
Ports Used by Cloudera Manager.....	41
Ports Used by Cloudera Navigator Key Trustee Server.....	46
Ports Used by Cloudera Runtime Components.....	46
Ports Used by DistCp.....	52
Ports Used by Third-Party Components.....	53
Runtime Cluster Hosts and Role Assignments.....	54
Allocating Hosts for Key Trustee Server and Key Trustee KMS.....	60
Service Dependencies in Cloudera Manager.....	61

Custom Installation Solutions.....62

Creating Virtual Images of Cluster Hosts.....	62
Creating a Pre-Deployed Cloudera Manager Host.....	62
Instantiating a Cloudera Manager Image.....	63
Creating a Pre-Deployed Worker Host.....	64
Instantiating a Worker Host.....	65
Configuring a Custom Java Home Location.....	65
Creating a Runtime Cluster Using a Cloudera Manager Template.....	66
Exporting the Cluster Configuration.....	66
Preparing a New Cluster.....	67
Creating the Template.....	67
Importing the Template to a New Cluster.....	71
Sample Python Code.....	72

Local Package and Parcel Repositories.....73

Understanding Package Management.....	73
Repository Configuration Files.....	73
Listing Repositories.....	74
Configuring a Local Package Repository.....	74
Creating a Permanent Internal Repository.....	74
Creating a Temporary Internal Repository.....	75
Configuring Hosts to Use the Internal Repository.....	76
Manually Install Cloudera Software Packages.....	76
Install Cloudera Manager Packages.....	76
Manually Install Cloudera Manager Agent Packages.....	76
Introduction to Parcels.....	77
Configuring a Local Parcel Repository.....	77
Using an Internally Hosted Remote Parcel Repository.....	78
Using a Local Parcel Repository.....	80

Production Installation: Installing Cloudera Manager, Cloudera Runtime, and Managed Services.....81

Step 1: Configure a Repository for Cloudera Manager.....	81
Step 2: Install Java Development Kit.....	82
Installing OpenJDK Using Cloudera Manager.....	82
Manually Installing OpenJDK.....	82
Manually Installing Oracle JDK.....	83
Step 3: Install Cloudera Manager Server.....	83
Install Cloudera Manager Packages.....	83
(Recommended) Enable Auto-TLS.....	84
Step 4: Install and Configure Databases.....	85
Required Databases.....	85
Install and Configure PostgreSQL for CDP.....	85
Installing PostgreSQL Server.....	86
Installing the psycopg2 Python Package.....	86
Configuring and Starting the PostgreSQL Server.....	86
Creating Databases for CDP.....	88
Step 5: Set up and configure the Cloudera Manager database.....	90
Syntax for scm_prepare_database.sh.....	90
Step 6: Install Runtime and Other Software.....	92
Installation Wizard.....	92

Step 7: Set Up a Cluster Using the Wizard.....	96
Select Services.....	96
Assign Roles.....	97
Setup Database.....	97
Enter Required Parameters.....	97
Review Changes.....	98
Command Details.....	98
Summary.....	98
Additional Steps for Apache Ranger.....	98
Installing Cloudera Navigator Key Trustee Server.....	99
Installing Key Trustee Server Using Cloudera Manager.....	99
Securing Key Trustee Server Host.....	100
Leveraging Native Processor Instruction Sets.....	100
Initializing Key Trustee Server.....	101
Installing Key Trustee KMS.....	102
Installing Key Trustee KMS Using Parcels.....	102
After You Install.....	102
Deploying Clients.....	103
Testing the Installation.....	103
Checking Host Heartbeats.....	103
Testing with Hue.....	103
Secure Your Cluster.....	104
Troubleshooting Installation Problems.....	104
Uninstalling Cloudera Manager and Managed Software.....	107
Record User Data Paths.....	107
Stop all Services.....	107
Deactivate and Remove Parcels.....	108
Delete the Cluster.....	108
Uninstall the Cloudera Manager Server.....	108
Uninstall Cloudera Manager Agent and Managed Software.....	109
Remove Cloudera Manager, User Data, and Databases.....	109
Uninstalling a Runtime Component From a Single Host.....	110

CDP Private Cloud Base Installation Guide

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

For non-production environments such as trial and proof-of-concept use cases, see the *Trial Installation* section for a simplified but limited installation procedure.

Related Information

[Trial Installation](#)

Version and Download Information

The following topics describe the version systems used by Cloudera Manager and Cloudera Runtime, as well as download information.

Cloudera Manager Version Information

- Cloudera Manager 7.0.3 is the current release of Cloudera Manager for CDP Private Cloud Base.

Cloudera Manager 7.0.3 Release Date: November 19, 2019

Cloudera Manager Download Information

Important: Access to Cloudera Manager binaries for production purposes requires authentication. To access the binaries at the locations below, you must first have an active subscription agreement and obtain a license key file along with the required authentication credentials (username and password).

The license key file and authentication credentials are provided in an email sent to customer accounts from Cloudera when a new license is issued. If you have an existing license with a CDP Private Cloud Base Edition entitlement, you might not have received an email. In this instance you can identify the authentication credentials from the license key file. If you do not have access to the license key, contact your account representative to receive a copy.

To identify your authentication credentials using your license key file, complete the following steps:

- From cloudera.com, log into the cloudera.com account associated with the CDP Private Cloud Base license and subscription agreement.
- On the [CDP Private Cloud Base Download page](#), click Download Now and scroll down to the Credential Generator.
- In the Generate Credentials text box, copy and paste the text of the “PGP Signed Message” within your license key file and click Get Credentials. The credentials generator returns your username and password.



Important: Take note of the authentication credentials. You might need them during installation to complete tasks such as configuring a remote parcel repository, or installing Cloudera Manager packages using a package manager such as YUM, APT, or other tools that you might be using in your environment.

When you obtain your authentication credentials, use them to form the URL where you can access the Cloudera Manager repository in the Cloudera Archive.

The repositories for Cloudera Manager 7.x are listed in the following tables:

Table 1: Cloudera Manager 7.0.3:

Repository Type	Repository Location	Repository File
RHEL 7 Compatible	https://[username]: [password]@archive.cloudera.com/p/ cm7/7.0.3/redhat7/yum	https://[username]: [password]@archive.cloudera.com/p/ cm7/7.0.3/redhat7/yum/cloudera-manager.repo

Cloudera Runtime Version Information

Cloudera Runtime is available in the following releases:

- Cloudera Runtime 7.0.3.1 is based on Apache Hadoop 3. For more information, see *Cloudera Runtime Component Versions*.

Cloudera Runtime 7.0.3.1 Release Date: November 19, 2019

Cloudera Runtime Download Information

Important: Access to Cloudera Runtime parcels for production purposes requires authentication. To access the parcels at the locations below, you must first have an active subscription agreement and obtain a license key file along with the required authentication credentials (username and password).

The license key file and authentication credentials are provided in an email sent to customer accounts from Cloudera when a new license is issued. If you have an existing license with a CDP Private Cloud Base Edition entitlement, you might not have received an email. In this instance you can identify the authentication credentials from the license key file. If you do not have access to the license key, contact your account representative to receive a copy.

To identify your authentication credentials using your license key file, complete the following steps:

- From cloudera.com, log into the cloudera.com account associated with the CDP Private Cloud Base license and subscription agreement.
- On the [CDP Private Cloud Base Download page](#), click Download Now and scroll down to the Credential Generator.
- In the Generate Credentials text box, copy and paste the text of the “PGP Signed Message” within your license key file and click Get Credentials. The credentials generator returns your username and password.

Important: Take note of the authentication credentials. You might need them during installation to complete tasks such as configuring a remote parcel repository.

When you obtain your authentication credentials, use them to form the URL where you can access the Runtime repository in the Cloudera Archive. Cloudera Manager can also download the Runtime parcels directly during the installation process.

The repositories for Cloudera Runtime 7.x are listed in the following tables:

Table 2: Cloudera Runtime 7.0.3.1:

Repository Type	Repository Location
Parcels	https://[username]:[password]@archive.cloudera.com/p/cdh7/7.0.3.0/ parcels

CDP Private Cloud Base Trial Download Information

You can try the CDP Private Cloud Base edition of Cloudera Data Platform for 60 days without obtaining a license key file.

To download CDP Private Cloud Base without obtaining a license key file, visit the [Cloudera Data Platform Data Center Trial Download](#) page, click Try Now, and follow the download instructions. When you install CDP Private Cloud Base without a license key, you are performing a trial installation that includes an embedded PostgreSQL database and is not suitable for a production environment. For more information on trial installations, see the trial installation documentation.

A 60-day trial of CDP Private Cloud Base Edition can be enabled permanently with the appropriate license. To obtain a CDP Private Cloud Base Edition license, fill in the [Contact Us](#) form or call 866-843-7207

Related Information

[Trial Installation](#)

Trial Installation

These topics provide instructions for installing the trial version of CDP Private Cloud Base in a non-production environment for demonstration and proof-of-concept use cases.

In these procedures, Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Server, an embedded PostgreSQL database, Cloudera Manager Agent, Cloudera Runtime, and other managed services on cluster hosts. Cloudera Manager also configures databases for the Cloudera Manager Server and Hive Metastore, and optionally for Cloudera Management Service roles.

This installation method is recommended for trial deployments, but is not supported for production deployments because it is not designed to scale. To use this method, server and cluster hosts must satisfy the following requirements:

- You must be able to log in to the Cloudera Manager Server host using the root user account or an account that has passwordless sudo privileges.
- The Cloudera Manager Server host must have uniform SSH access on the same port to all hosts. For more information, see [Runtime and Cloudera Manager Networking and Security Requirements](#).
- All hosts must have access to standard package repositories for the operating system and either archive.cloudera.com or a local repository with the required installation files.
- SELinux must be disabled or set to permissive mode before running the installer.

Related Information

[CDP Private Cloud Base Trial Download Information](#)

[CDP Private Cloud Base Installation Guide](#)

Before You Begin a Trial Installation

Before you begin a trial installation, you must disable SELinux if you want the Cloudera Manager installer to run. You can also optionally configure an HTTP proxy.

(Optional) Configure an HTTP Proxy

The Cloudera Manager installer accesses archive.cloudera.com by using yum on RHEL systems. If your hosts access the Internet through an HTTP proxy, you can configure yum system-wide, to access archive.cloudera.com through a proxy.

To do so, modify the system configuration on every cluster host as follows:

OS	File	Property
RHEL-compatible	/etc/yum.conf	proxy=http://server:port/

Disable SELinux



Note: CDP Private Cloud Base is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in enforcing mode. Cloudera is not responsible for SELinux policy development, support, or enforcement. If you experience issues running Cloudera software with SELinux enabled, contact your OS provider for assistance.

If you are using SELinux in enforcing mode, Cloudera Support can request that you disable SELinux or change the mode to permissive to rule out SELinux as a factor when investigating reported issues.

Although Cloudera supports running Cloudera software with SELinux enabled, the Cloudera Manager installer will not proceed if SELinux is enabled. Disable SELinux or set it to permissive mode before running the installer.

After you have installed and deployed Cloudera Manager and Runtime, you can re-enable SELinux by changing SELINUX=permissive back to SELINUX=enforcing in `/etc/selinux/config` (or `/etc/sysconfig/selinux`), and then running the following command to immediately switch to enforcing mode:

```
setenforce 1
```

If you are having trouble getting Cloudera Software working with SELinux, contact your OS vendor for support. Cloudera is not responsible for developing or supporting SELinux policies.

Installing a Trial Cluster

In this procedure, Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Server, embedded PostgreSQL database, Cloudera Manager Agent, Runtime, and managed service software on cluster hosts. Cloudera Manager also configures databases for the Cloudera Manager Server and Hive Metastore and optionally for Cloudera Management Service roles.



Important: This procedure is intended for trial and proof-of-concept deployments only. It is not supported for production deployments because it is not designed to scale.

Cluster Host Requirements:

The hosts you intend to use must satisfy the following requirements:

- You must be able to log in to the Cloudera Manager Server host using the root user account or an account that has passwordless sudo privileges.
- The Cloudera Manager Server host must have uniform SSH access on the same port to all hosts. For more information, see *Runtime and Cloudera Manager Networking and Security Requirements*.
- All hosts must have access to standard package repositories for the operating system and either `archive.cloudera.com` or a local repository with the required installation files.
- SELinux must be disabled or set to permissive mode before running the installer.

Refer to the following topics for the steps required to install a trial cluster.

Step 1: Run the Cloudera Manager Server Installer

Download the Cloudera Manager installer to the cluster host to which you are installing the Cloudera Manager Server. By default, the automated installer binary (`cloudera-manager-installer.bin`) installs the highest version of Cloudera Manager.

Before you begin

For information on downloading the CDP Private Cloud Base Trial installer, see [CDP Private Cloud Base Trial Download Information](#) on page 7.

Procedure

1. Run the Cloudera Manager installer:

- a) Change cloudera-manager-installer.bin to have execute permissions:

```
chmod u+x cloudera-manager-installer.bin
```

- b) Run the Cloudera Manager Server installer:

```
sudo ./cloudera-manager-installer.bin
```

- c) For clusters without Internet access: Install Cloudera Manager packages from a local repository:

```
sudo ./cloudera-manager-installer.bin --skip_repo_package=1
```

2. Read and accept the associated license agreements. After you accept the licenses, the installer does the following:

- a. Installs the Cloudera Manager repository files.
- b. Installs the Oracle JDK.
- c. Installs the Cloudera Manager Server and embedded PostgreSQL packages.
- d. Starts the embedded PostgreSQL database and Cloudera Manager Server.



Note: If the installation is interrupted, run the following command on the Cloudera Manager Server host before you retry the installation:

```
sudo /usr/share/cmfd/uninstall-cloudera-manager.sh
```

Log files for the installer are stored in /var/log/cloudera-manager-installer/.

3. Exit the installer:

- a) When the installation completes, the complete URL for the Cloudera Manager Admin Console displays, including the port number (7180 by default). Make a note of this URL.
- b) Press Enter to choose OK to exit the installer, and then again to acknowledge the successful installation.
- c) Wait several minutes for the Cloudera Manager Server to start. To observe the startup process, run `sudo tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log` on the Cloudera Manager Server host. When you see the following 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*.

What to do next

If you plan to use Ranger, you must create and configure the Ranger database before you proceed with cluster configuration.

Step 2. Configure the Ranger and DAS Databases

If you plan to use Data Analytics Studio or Ranger in a trial cluster, you must create DAS and Ranger databases.

About this task

If you plan to use Data Analytics Studio in your trial cluster, you must create an external PostgreSQL 9.6 database. This is because the embedded database created by the trial installer is PostgreSQL version 9.2, which is incompatible with DAS. For more information on creating a DAS database, see the topics *Required Databases* and *Creating Databases for CDP*.

If you plan to use Ranger during the trial, you must create a Ranger database. A script is provided to create the database.

Complete this task before you setup a cluster with Ranger enabled:

Procedure

1. To create the Ranger database, run the `gen_embedded_ranger_db.sh` located in `/opt/cloudera/cm/bin` on the Cloudera Manager host.
The output of the script gives you the default Ranger database host, default database name, default database user, and the Ranger database user password to use during cluster setup.
2. Take note of the script output. Although the database name and database user are default values, you will need to enter the host and Ranger database user password on the Enter Required Parameters page when you configure a cluster.

What to do next

The wizard that follows will deploy a cluster and start the Ranger service. At that point, the Ranger service enables the necessary plugins. You should verify that the Ranger plugins for HDFS and Solr are enabled. For more information see *Additional Steps for Apache Ranger*.

If cluster deployment fails, be sure to click Resume in the wizard after you fix any issues. If you do not click Resume, the Ranger service will not enable all of the necessary plugins.

Related Information

[Required Databases](#)

[Creating Databases for CDP](#)

Step 3: Install Runtime Using the Wizard

Proceed through the installation wizard to specify hosts, install and configure Cloudera Runtime, and more.

Log Into the Cloudera Manager Console

1. 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.
2. 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.

Upload License File

On the Upload License File page, you can select either the trial version of CDP Data Center or upload a license file:

1. Choose one of the following options:
 - Upload Cloudera Data Platform License
 - Try Cloudera Data Platform for 60 days. The CDP Data Center trial does not require a license file, but the trial expires after 60 days.
2. If you choose the CDP Data Center Edition Trial, you can upload a license file at a later time. Read the license agreement and click the checkbox labeled Yes, I accept the Cloudera Standard License Terms and Conditions if you accept the terms and conditions of the license agreement. Then click Continue.
3. If you have a license file for CDP Data Center, upload the license file:
 - a. Select Upload Cloudera Data Platform License.
 - b. Click Upload License File.
 - c. Browse to the location of the license file, select the file, and click Open.
 - d. Click Upload.
 - e. Click Continue.

4. Information is displayed indicating what the Runtime installation includes. At this point, you can click the Support drop-down menu to access online Help or the Support Portal.
5. Click Continue to proceed with the installation.

Welcome (Add Cluster - Installation)

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

Click Continue to proceed with the installation.

Cluster Basics

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

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

For new installations, Regular Cluster is the only option. You cannot add a compute cluster if you do not have an existing base cluster.

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

Enter a cluster name and then click Continue.

Specify Hosts

Choose which hosts will run Runtime and other managed services.



Note: If you have enabled Auto-TLS, you must include the Cloudera Manager server host when you specify hosts.

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

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



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

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

2. Click Search. If there are a large number of hosts on your cluster, wait a few moments to allow them to be discovered and shown in the wizard. If the search is taking too long, you can stop the scan by clicking Abort

Scan. You can modify the search pattern and repeat the search as many times as you need until you see all of the expected hosts.



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

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

The Select Repository screen displays.

Select Repository

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

In the Cloudera Manager Agent section:

1. Select either Public Cloudera Repository or Custom Repository for the Cloudera Manager Agent software.
2. If you select Custom Repository, do not include the operating system-specific paths in the URL. For instructions on setting up a custom repository, see *Configuring a Local Package Repository*.
3. Select the installation method for Runtime and other software. For this trial cluster, select Use Parcels (Recommended).
4. Select the version of Runtime to install. If 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 Runtime 7 are documented in *Cloudera Runtime 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.



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

5. Specify any Additional Parcels you want to install.
6. Click Continue.

The Select JDK page displays.

Select JDK



Note: CDP Data Center is no longer bundled with Oracle JDK software. Cloudera provides a supported version of OpenJDK.

If you installed your own JDK version, such as Oracle JDK 8, in *Step 2: Install Java Development Kit*, select Manually manage JDK.

To allow Cloudera Manager to automatically install the OpenJDK on cluster hosts, select Install a Cloudera-provided version of OpenJDK.

To install the default OpenJDK that is provided by your operating system, select Install a system-provided version of OpenJDK.

After checking the applicable boxes, click Continue.

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.

After installing the Cloudera Manager Agent on all hosts, click Continue. The Install Parcels page displays.

Install Parcels

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 page displays.

Inspect Cluster

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

1. Run the Inspect Network Performance tool. You can click Advanced Options to customize some ping parameters.
2. After the network inspector completes, click Show Inspector Results to view the results in a new tab.
3. Address any reported issues, and click Run Again (if applicable).
4. Click Inspect Hosts to run the Host Inspector utility.
5. After the host inspector completes, click Show Inspector Results to view the results in a new tab.
6. Address any reported issues, and click Run Again (if applicable).

If the reported issues cannot be resolved in a timely manner, and you want to abandon the cluster creation wizard to address them, select the radio button labeled Quit the wizard and Cloudera Manager will delete the temporarily created cluster and then click Continue.

Otherwise, after addressing any identified problems, select the radio button labeled I understand the risks, let me continue with cluster creation, and then click Continue.

This completes the Cluster Installation wizard and launches the Add Cluster - Configuration wizard.

Step 4: 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:

Data Engineering

HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Zeppelin, Livy, and Hue

Data Mart

HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Impala, and Hue

Data Warehouse

HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on LLAP, and Data Analytics Studio

Operational Database

HDFS, Ranger, Atlas, and HBASE

Custom Services

Choose your own services. Services required by chosen services will automatically be included.

For compute clusters, you can choose from:

Data Engineering

Spark, Oozie, Hive on Tez, Data Analytics Studio, HDFS, YARN, and YARN Queue Manager

Spark

Spark, Oozie, YARN, and YARN Queue Manager

Data Mart

Impala

Custom Services

Choose your own services. Services required by chosen services will automatically be included.

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*.

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

Setup Database

When using the Cloudera Manager installer with the embedded database, the Setup Database page is pre-populated with the database names and passwords. 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. Some databases will be created in a future step. For these, the words Skipped. Cloudera Manager will create this database in a later step. appear next to the green checkmark.

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

Enter Required Parameters

The **Enter Required Parameters** page lists required parameters for DAS, the Cloudera Manager API client, and Ranger.

The DAS database hostname, database name, database username, and database password were configured when you created the required DAS database. The default database name is “das” and the default database user is “das”.

If you do not have an existing user for the Cloudera Manager API client, use the default username and password “admin” for both the The Existing Cloudera Manager API Client Username and The Existing Cloudera Manager API Client Password.

The Ranger Admin user, Usersync user, Tagsync User, and KMS Keyadmin User are created during cluster deployment. In this page you must give a password for each of these users.



Note: Passwords for the Ranger Admin, Usersync, Tagsync, and KMS Keyadmin users must be a minimum of 8 characters long, with at least one alphabetic and one numeric character. The following characters are not valid: " ' \ ` ' .

The Ranger database host, name, user, and user password were configured when you created the required Ranger database. If you ran the `gen_embedded_ranger_db.sh` script to create the Ranger database, the output of the script contained the host and database user password. Enter those here. The default database name is “ranger” and the default database user is “rangeradmin.”

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 Command Details page displays.

Command Details

The Command Details 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 to go to the Summary page.

Summary

The Summary page reports the success or failure 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.

Stopping the Embedded PostgreSQL Database

To stop the embedded PostgreSQL database, stop the services and servers in the order listed below.

Procedure

1. Log into the Cloudera Manager user interface and stop the services that have a dependency on the Hive metastore (Hue, Impala, and Hive) in the following order:
 - Stop the Hue and Impala services.
 - Stop the Hive service.
2. On the Cloudera Manager **Home** page, click the 3 vertical dots next to Cloudera Management Service and select Stop to stop the Cloudera Management Service.

3. Stop the Cloudera Manager Server.

RHEL 7:

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

4. Stop the Cloudera Manager Server database.

RHEL 7:

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

Starting the Embedded PostgreSQL Database

To start the embedded PostgreSQL database, start the servers and services in the order listed below.

Procedure

1. Start the Cloudera Manager Server database.

RHEL 7:

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

2. Start the Cloudera Manager Server.

RHEL 7:

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

3. Log into Cloudera Manager and start the Cloudera Manager Service. On the Cloudera Manager **Home** page, click the 3 vertical dots next to Cloudera Management Service and select Start.
4. In the Cloudera Manager user interface, start the services that have a dependency on the Hive metastore (Hue, Impala, and Hive) in the following order:
 - Start the Hive service.
 - Start the Hue and Impala services.

Changing Embedded PostgreSQL Database Passwords

The embedded PostgreSQL database has generated user accounts and passwords. You can change a password associated PostgreSQL database account.

About this task

You can see the generated accounts and passwords during the installation process and you should record them at that time.

To find information about the PostgreSQL database account that the Cloudera Manager Server uses, read the `/etc/cloudera-scm-server/db.properties` file:

```
# cat /etc/cloudera-scm-server/db.properties

Auto-generated by scm_prepare_database.sh
#
Sat Oct 1 12:19:15 PDT 201
#
com.cloudera.cmf.db.type=postgresql
com.cloudera.cmf.db.host=localhost:7432
com.cloudera.cmf.db.name=scm
```

```
com.cloudera.cmf.db.user=scm
com.cloudera.cmf.db.password=TXqEESuhj5
```

To change a password associated with a PostgreSQL database account:

Procedure

1. Obtain the root password from the `/var/lib/cloudera-scm-server-db/data/generated_password.txt` file:

```
# cat /var/lib/cloudera-scm-server-db/data/generated_password.txt

MnPwGeWaip

The password above was generated by /usr/share/cmf/bin/initialize_embedded_db.sh (part of the cloudera-scm-server-db package)
and is the password for the user 'cloudera-scm' for the database in the
current directory.

Generated at Fri Jun 29 16:25:43 PDT 2012.
```

2. On the host on which the Cloudera Manager Server is running, log into PostgreSQL as the root user:

```
psql -U cloudera-scm -p 7432 -h localhost -d postgres
Password for user cloudera-scm: MnPwGeWaip
psql (8.4.18)
Type "help" for help.

postgres=#
```

3. Determine the database and owner names:

```
postgres=# \l

               List of databases
   Name      |  Owner   | Encoding | Collation |  Ctype  |
-----+-----+-----+-----+-----+
 amon        | amon     | UTF8     | en_US.UTF8 | en_US.UTF8 |
 hive        | hive     | UTF8     | en_US.UTF8 | en_US.UTF8 |
 nav         | nav      | UTF8     | en_US.UTF8 | en_US.UTF8 |
 navms       | navms    | UTF8     | en_US.UTF8 | en_US.UTF8 |
 postgres    | cloudera-scm | UTF8     | en_US.UTF8 | en_US.UTF8 |
 rman        | rman     | UTF8     | en_US.UTF8 | en_US.UTF8 |
 scm         | scm      | UTF8     | en_US.UTF8 | en_US.UTF8 |
 template0   | cloudera-scm | UTF8     | en_US.UTF8 | en_US.UTF8 |
 dera-scm"   |           |           |           |           |
                                     : "cloudera
-scm"=CTc/"cloudera-scm"
 template1   | cloudera-scm | UTF8     | en_US.UTF8 | en_US.UTF8 |
 oudera-scm" |           |           |           |           |
                                     : "cloude
ra-scm"=CTc/"cloudera-scm"
(9 rows)
```

4. Set the password for an owner using the `\password` command. For example, to set the password for the amon owner, do the following:

```
postgres=# \password amon
Enter new password:
Enter it again:
```

5. Configure the role with the new password:
 - a) In the Cloudera Manager Admin Console, select ClustersCloudera Management Service.
 - b) Click the Configuration tab.
 - c) In the Scope section, select the role where you are configuring the database.
 - d) Select CategoryDatabase category.
 - e) Set the *Role Name* Database Password property.
 - f) Enter a Reason for change, and then click Save Changes to commit the changes.

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

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

Cloudera Manager provides an embedded PostgreSQL database server for trial and proof of concept deployments when creating a cluster. To remind users that this embedded database is not suitable for production, Cloudera Manager displays the banner text: "You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database before moving into production."

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



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

Prerequisites

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

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



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

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

- All health issues with your cluster have been resolved.

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



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

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

Identify Roles that Use the Embedded Database Server

Before you can migrate to another database server, you must first identify the databases using the embedded database server.

About this task

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

To identify which roles are using the Cloudera Manager embedded database server:

Procedure

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

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

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

Table 3: Cloudera Manager Embedded Database Server Databases

Service	Role	Default Database Name	Default Username
Cloudera Manager Server		scm	scm
Cloudera Management Service	Activity Monitor	amon	amon
Hive	Hive Metastore Server	hive	hive
Hue	Hue Server	hue	7uu7uu7uhue
Oozie	Oozie Server	oozie_oozie_server	oozie_oozie_server
Cloudera Management Service	Reports Manager	rman	rman

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

For Cloudera Management Services:

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



Note:

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

- Activity Monitor
- Reports Manager

For the Oozie Service:

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

For Hive and Hue Services:

1. Select the specific service > Configuration, and type "database host" in the Search field.
 2. Confirm that the hostname is the Cloudera Manager Server.
 3. In the Search field, type "database port" and confirm that the port is 7432.
 4. Repeat these steps for each of the services (Hive and Hue).
4. Verify the database names in the embedded database server match the database names on your list (Step 2). Databases that exist on the database server and not used by their roles do not need to be migrated. This step is to confirm that your list is correct.



Note: Do not add the postgres, template0, or template1 databases to your list. These are used only by the PostgreSQL server.

```
psql -h localhost -p 7432 -U cloudera-scm -l
```

```
Password for user cloudera-scm: <password>
```

		List of databases			
Name	Access	Owner	Encoding	Collate	Ctype
-----+-----+-----+-----+-----+-----					
amon		amon	UTF8	en_US.UTF8	en_US.U
TF8					
hive		hive	UTF8	en_US.UTF8	en_US.UT
F8					
hue		hue	UTF8	en_US.UTF8	en_US
.UTF8					
navms		navms	UTF8	en_US.UTF8	en_US.
UTF8					
oozie_oozie_server		oozie_oozie_server	UTF8	en_US.UTF8	en_US.U
TF8					
postgres		cloudera-scm	UTF8	en_US.UTF8	en_US.UT
F8					
rman		rman	UTF8	en_US.UTF8	en_US
.UTF8					

```
scm | scm | UTF8 | en_US.UTF8 | en_US.
UTF8 |
template0 | cloudera-scm | UTF8 | en_US.UTF8 | en_US.U
TF8 | =c/"cloudera-scm"
template1 | cloudera-scm | UTF8 | en_US.UTF8 | en_US.
UTF8 | =c/"cloudera-scm"
(12 rows)
```

Results

You should now have a list of all roles and database names that use the embedded database server, and are ready to proceed with the migration of databases from the embedded database server to the external PostgreSQL database server.

Migrate Databases from the Embedded Database Server to the External PostgreSQL Database Server

After you identify the roles that use the embedded database, you are ready to migrate from the embedded database server to an external PostgreSQL database server.

About this task

While performing this procedure, ensure that the Cloudera Manager Agents remain running on all hosts. Unless otherwise specified, when prompted for a password use the cloudera-scm password.



Note: After completing this migration, you cannot delete the cloudera-scm postgres superuser unless you remove the access privileges for the migrated databases. Minimally, you should change the cloudera-scm postgres superuser password.

Procedure

1. In Cloudera Manager, stop the cluster services identified as using the embedded database server. Be sure to stop the Cloudera Management Service as well. Also be sure to stop any services with dependencies on these services. The remaining Runtime services will continue to run without downtime.



Note: If you do not stop the services from within Cloudera Manager before stopping Cloudera Manager Server from the command line, they will continue to run and maintain a network connection to the embedded database server. If this occurs, then the embedded database server will ignore any command line stop commands (Step 2) and require that you manually kill the process, which in turn causes the services to crash instead of stopping cleanly.

2. Navigate to Hosts > All Hosts, and make note of the number of roles assigned to hosts. Also take note whether or not they are in a commissioned state. You will need this information later to validate that your scm database was migrated correctly.
3. Stop the Cloudera Manager Server. To stop the server:

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

4. Obtain and save the embedded database superuser password (you will need this password in subsequent steps) from the generated_password.txt file:

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

5. Export the PostgreSQL user roles from the embedded database server to ensure the correct users, permissions, and passwords are preserved for database access. Passwords are exported as an md5sum and are not visible in plain text. To export the database user roles (you will need the cloudera-scm user password):

```
pg_dumpall -h localhost -p 7432 -U cloudera-scm -v --roles-only -f "/var/
tmp/cloudera_user_roles.sql"
```

6. Edit the `/var/tmp/cloudera_user_roles.sql` file to remove any `CREATE ROLE` and `ALTER ROLE` commands for databases not in your list. Leave the entries for the cloudera-scm user untouched, because this user role is used during the database import.



Important: If the external PostgreSQL database is an Amazon's Relational Database Service (RDS), then remove all entries for `ALTER ROLE` or `CREATE ROLE` commands from the `/var/tmp/cloudera_user_roles.sql` file for the Cloudera Manager database's user such as cloudera-scm, and then add the following command for the same user:

```
CREATE ROLE cloudera-scm WITH NOSUPERUSER INHERIT NOCREATEROLE NOCRE
ATEDB LOGIN NOREPLICATION NOBYPASSRLS PASSWORD '<stripped>';
```

7. Export the data from each of the databases on your list you created in *Identify Roles that Use the Embedded Database Server*:

```
pg_dump -F c -h localhost -p 7432 -U cloudera-scm [database_name] > /var/
tmp/[database_name]_db_backup-$(date +%m-%d-%Y).dump
```

The following is a sample data export command for the scm database:

```
pg_dump -F c -h localhost -p 7432 -U cloudera-scm scm > /var/tmp/scm_db_
backup-$(date +%m-%d-%Y).dump
```

Password:

8. Stop and disable the embedded database server:

```
service cloudera-scm-server-db stop
chkconfig cloudera-scm-server-db off
```

Confirm that the embedded database server is stopped:

```
netstat -at | grep 7432
```

9. Back up the Cloudera Manager Server database configuration file:

```
cp /etc/cloudera-scm-server/db.properties /etc/cloudera-scm-server/db.pr
operties.embedded
```

10. Copy the file `/var/tmp/cloudera_user_roles.sql` and the database dump files from the embedded database server host to `/var/tmp` on the external database server host:

```
cd /var/tmp
scp cloudera_user_roles.sql *.dump <user>@<postgres-server>:/var/tmp
```

11. Import the PostgreSQL user roles into the external database server.

The external PostgreSQL database server superuser password is required to import the user roles. If the superuser role has been changed, you will be prompted for the username and password.



Note: Only run the command that applies to your context; do not execute both commands.

- To import users when using the default PostgreSQL superuser role:

```
sudo -u postgres psql -f /var/tmp/cloudera_user_roles.sql
```

- To import users when the superuser role has been changed:

```
psql -h <database-hostname> -p <database-port> -U <superuser> -f /var/tmp/cloudera_user_roles.sql
```

For example:

```
psql -h pg-server.example.com -p 5432 -U postgres -f /var/tmp/cloudera_user_roles.sql
```

```
Password for user postgres
```

12. Import the Cloudera Manager database on the external server. First copy the database dump files from the Cloudera Manager Server host to your external PostgreSQL database server, and then import the database data:



Note: To successfully run the `pg_restore` command, there must be an existing database on the database server to complete the connection; the existing database will not be modified. If the `-d <existing-database>` option is not included, then the `pg_restore` command will fail.

```
pg_restore -C -h <database-hostname> -p <database-port> -d <existing-database> -U cloudera-scm -v <data-file>
```

Repeat this import for each database.

The following example is for the scm database:

```
pg_restore -C -h pg-server.example.com -p 5432 -d postgres -U cloudera-scm -v /var/tmp/scm_server_db_backup-20180312.dump
```

```
pg_restore: connecting to database for restore
Password:
```

13. Update the Cloudera Manager Server database configuration file to use the external database server. Edit the `/etc/cloudera-scm-server/db.properties` file as follows:

- Update the `com.cloudera.cmf.db.host` value with the hostname and port number of the external database server.
- Change the `com.cloudera.cmf.db.setupType` value from "EMBEDDED" to "EXTERNAL".

14. Start the Cloudera Manager Server and confirm it is working:

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

Note that if you start the Cloudera Manager GUI at this point, it may take up to five minutes after executing the start command before it becomes available.

In Cloudera Manager Server, navigate to Hosts > All Hosts and confirm the number of roles assigned to hosts (this number should match what you found in Step 2); also confirm that they are in a commissioned state that matches what you observed in Step 2.

15. Update the role configurations to use the external database hostname and port number. Only perform this task for services where the database has been migrated.

For Cloudera Management Services:

- a. Select Cloudera Management Service > Configuration, and type "7432" in the Search field.
- b. Change any database hostname properties from the embedded database to the external database hostname and port number.
- c. Click Save Changes.

For the Oozie Service:

- a. Select Oozie service > Configuration, and type "7432" in the Search field.
- b. Change any database hostname properties from the embedded database to the external database hostname and port number.
- c. Click Save Changes.

For Hive and Hue Services:

- a. Select the specific service > Configuration, and type "database host" in the Search field.
- b. Change the hostname from the embedded database name to the external database hostname.
- c. Click Save Changes.

16. Start the Cloudera Management Service and confirm that all management services are up and no health tests are failing.

17. Start all Services via the Cloudera Manager web UI. This should start all services that were stopped for the database migration. Confirm that all services are up and no health tests are failing.

18. On the embedded database server host, remove the embedded PostgreSQL database server:

- a) Make a backup of the /var/lib/cloudera-scm-server-db/data directory:

```
tar czvf /var/tmp/embedded_db_data_backup-$(date +"%m-%d-%Y").tgz /var/lib/cloudera-scm-server-db/data
```

- b) Remove the embedded database package:

For RHEL/SLES:

```
rpm --erase cloudera-manager-server-db-2
```

```
apt-get remove cloudera-manager-server-db-2
```

- c) Delete the /var/lib/cloudera-scm-server-db/data directory.

Production Installation: Before You Install

Before you begin a production installation of Cloudera Manager, Cloudera Runtime, and other managed services, review the Cloudera Data Platform 7 Requirements and Supported Versions, in addition to the Cloudera Data Platform Release Notes.

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

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.

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

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. 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.

Table 4: Cloudera Manager Server

Configuration Topic	Cloudera Manager Server Configuration
Default Storage Location	<p>RDBMS:</p> <p>Any Supported RDBMS.</p> <p>Disk:</p> <p>Cloudera Manager Server Local Data Storage Directory (<code>command_storage_path</code>) 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.</p> <p>Default setting: <code>/var/lib/cloudera-scm-server/</code></p>
Storage Configuration Defaults, Minimum, or Maximum	There are no direct storage defaults relevant to this entity.

Configuration Topic	Cloudera Manager Server Configuration
Where to Control Data Retention or Size	<p>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 AdministrationSettings and edit the following property:</p> <p>Command Eviction Age</p> <p>Length of time after which inactive commands are evicted from the database.</p> <p>Default is two years.</p>
Sizing, Planning & Best Practices	<p>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.</p> <p>Make sure that you perform regular, verified, remotely-stored backups of the Cloudera Manager Server database.</p>

Table 5: Cloudera Management Service - Activity Monitor Configuration

Configuration Topic	Activity Monitor
Default Storage Location	Any Supported RDBMS.
Storage Configuration Defaults / Minimum / Maximum	Default: 14 Days worth of MapReduce (MRv1) jobs/tasks
Where to Control Data Retention or Size	<p>You control Activity Monitor storage usage by configuring the number of days or hours of data to retain. Older data is purged.</p> <p>To configure data retention in the Cloudera Manager Administration Console:</p> <ol style="list-style-type: none"> 1. Go to the Cloudera Management Service. 2. Click the Configuration tab. 3. Select ScopeActivity Monitor or Cloudera Management Service (Service-Wide). 4. Select CategoryMain. 5. Locate the following properties or search for them by typing the property name in the Search box: <ul style="list-style-type: none"> Purge Activities Data at This Age <p>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).</p> Purge Attempts Data at This Age <p>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).</p> Purge MapReduce Service Data at This Age <p>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).</p> 6. Enter a Reason for change, and then click Save Changes to commit the changes.

Configuration Topic	Activity Monitor
Sizing, Planning, and Best Practices	<p>The Activity Monitor only monitors MapReduce jobs, and does not monitor YARN applications.</p> <p>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.</p> <p>For example, consider the following test cluster and usage:</p> <ul style="list-style-type: none"> • A simulated 1000-host cluster, each host with 32 slots • MapReduce jobs with 200 attempts (tasks) per activity (job) <p>Sizing observations for this cluster:</p> <ul style="list-style-type: none"> • 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 6: 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	<ul style="list-style-type: none"> • 10 GiB Services Time Series Storage • 1 GiB Impala Query Storage • 1 GiB YARN Application Storage <p>Total: ~12 GiB Minimum (No Maximum)</p>

Configuration Topic	Service Monitor Configuration
Where to Control Data Retention or Size	<p>Service Monitor data growth is controlled by configuring the maximum amount of storage space it can use.</p> <p>To configure data retention in Cloudera Manager Administration Console:</p> <ol style="list-style-type: none"> 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 <i>propertyName</i> property or search for it by typing its name in the Search box. <p>Time-Series Storage</p> <p>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.</p> <p>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.</p> <p>Select Cloudera Management ServiceCharts 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.</p> <p>Impala Storage</p> <p>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.</p> <p>YARN Storage</p> <p>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.</p> <ol style="list-style-type: none"> 6. Enter a Reason for change, and then click Save Changes to commit the changes.
Sizing, Planning, and Best Practices	<p>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.</p>

Table 7: 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

Configuration Topic	Host Monitor Configuration
Where to Control Data Retention or Size	<p>Host Monitor data growth is controlled by configuring the maximum amount of storage space it can use.</p> <p>See <i>Data Storage for Monitoring Data</i>.</p> <p>To configure these data retention configuration properties in the Cloudera Manager Administration Console:</p> <ol style="list-style-type: none"> 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. <p>Time-Series Storage</p> <p>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.</p> <p>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.</p> <p>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.</p> <ol style="list-style-type: none"> 6. Enter a Reason for change, and then click Save Changes to commit the changes.
Sizing, Planning and Best Practices	<p>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.</p>

Table 8: 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.
Storage Configuration Defaults	5,000,000 events retained
Where to Control Data Retention or Minimum /Maximum	<p>The amount of storage space the Event Server uses is influenced by configuring how many discrete events it can retain.</p> <p>To configure data retention in Cloudera Manager Administration Console,</p> <ol style="list-style-type: none"> 1. Go the Cloudera Management Service. 2. Click the Configuration tab. 3. Select Scope Event Server or Cloudera Management Service (Service-Wide). 4. Select CategoryMain. 5. Edit the following property: Maximum Number of Events in the Event Server Store <p>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</p> <ol style="list-style-type: none"> 6. Enter a Reason for change, and then click Save Changes to commit the changes.


Configuration Topic	Event Server Configuration
Sizing, Planning, and Best Practices	<p>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.</p> <p> 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.</p>

Table 9: Cloudera Management Service - Reports Manager

Configuration Topic	Reports Manager Configuration
Default Storage Location	<p>RDBMS:</p> <p>Any Supported RDBMS.</p> <p>Disk:</p> <p>/var/lib/cloudera-scm-headlamp/ on the host where the Reports Manager role is configured to run.</p>
Storage Configuration Defaults	<p>RDBMS:</p> <p>There are no configurable parameters to directly control the size of this data set.</p> <p>Disk:</p> <p>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, but also on the HDFS file path complexity. Longer file paths contribute to more space utilization.</p>
Where to Control Data Retention or Minimum / Maximum	<p>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.</p>
Sizing, Planning, and Best Practices	<p>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.</p> <p>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.</p>

Table 10: Cloudera Navigator - Navigator Audit Server

Configuration Topic	Navigator Audit Server Configuration
Default Storage Location	Any Supported RDBMS.
Storage Configuration Defaults	Default: 90 Days retention

Configuration Topic	Navigator Audit Server Configuration
Where to Control Data Retention or Min/Max	<p>Navigator Audit Server storage usage is controlled by configuring how many days of data it can retain. Any older data is purged.</p> <p>To configure data retention in the Cloudera Manager Administration Console:</p> <ol style="list-style-type: none"> 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 CategoryMain. 5. Locate the Navigator Audit Server Data Expiration Period property or search for it by typing its name in the Search box. <p>Navigator Audit Server Data Expiration Period</p> <p>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.</p> <ol style="list-style-type: none"> 6. Click Save Changes to commit the changes.
Sizing, Planning, and Best Practices	<p>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).</p> <p>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 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.</p> <p>Notes:</p> <ul style="list-style-type: none"> • 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. <p>To map Cloudera Navigator versions to Cloudera Manager versions, see <i>Product Compatibility Matrix for Cloudera Navigator</i>.</p>

Table 11: Cloudera Navigator - Navigator Metadata Server

Configuration Topic	Navigator Metadata Server Configuration
Default Storage Location	<p>RDBMS:</p> <p>Any Supported RDBMS.</p> <p>Disk:</p> <p>/var/lib/cloudera-scm-navigator/ on the host where the Navigator Metadata Server role is configured to run.</p>
Storage Configuration Defaults	<p>RDBMS:</p> <p>There are no exposed defaults or configurations to directly cull or purge the size of this data set.</p> <p>Disk:</p> <p>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).</p>

Configuration Topic	Navigator Metadata Server Configuration
Where to Control Data Retention or Min/Max	<p>RDBMS:</p> <p>The Navigator Metadata Server database should be carefully tuned to support large volumes of metadata.</p> <p>Disk:</p> <p>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.</p>
Sizing, Planning, and Best Practices	<p>Memory:</p> <p><i>See Navigator Metadata Server Tuning.</i></p> <p>RDBMS:</p> <p>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.</p> <p>Use the product compatibility matrix to map Cloudera Navigator and Cloudera Manager versions.</p> <p>Disk:</p> <p>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.</p>

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 Cloudera Runtime and other components require adequate disk space in the following locations:

Table 12: Parcel Lifecycle Management

Parcel Lifecycle Path (default)	Notes
Local Parcel Repository Path (/opt/cloudera/parcel-repo)	<p>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.</p> <p>Sizing and Planning</p> <p>The default location is /opt/cloudera/parcel-repo but you can configure another local filesystem location on the host where Cloudera Manager Server runs.</p> <p>Provide sufficient space to hold all the parcels you download from all configured Remote Parcel Repository URLs. Cloudera Manager deployments that manage multiple clusters store all applicable parcels for all clusters.</p> <p>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.</p> <p>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.</p> <p>Lifecycle Management and Best Practices</p> <p>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.</p> <p>Backup Considerations</p> <p>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.</p>
Parcel Cache (/opt/cloudera/parcel-cache)	<p>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.</p> <p>Sizing and Planning</p> <p>Provide sufficient space per-host to hold all the parcels you distribute to each host.</p> <p>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.</p> <p>To configure this behavior in the Cloudera Manager Administration Console, select AdministrationSettingsParcelsRetain Downloaded Parcel Files</p>

Parcel Lifecycle Path (default)	Notes
Host Parcel Directory (/opt/cloudera/parcels)	<p>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.</p> <p>Sizing and Planning</p> <p>Provide sufficient space on each host to hold all the parcels you distribute to each host. Be aware that the typical Runtime or 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.</p> <p>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.</p> <p>To configure this behavior in the Cloudera Manager Administration Console, select AdministrationSettingsParcels and configure the following property:</p> <p>Automatically Remove Old Parcels</p> <p>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.</p> <p>The default value is Disabled.</p> <p>Number of Old Parcel Versions to Retain</p> <p>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.</p> <p>The default value is 3.</p>

Table 13: Management Service Lifecycle - Space Reclamation Tasks

Task	Description
Activity Monitor (One-time)	<p>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.</p>
Service Monitor and Host Monitor (One-time)	<p>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.</p> <p>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.</p>
Ongoing Space Reclamation	<p>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.</p>

Log File Storage Space

All 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.

Configure Network Names

You must configure each host in the cluster to ensure that all members can communicate with each other.

About this task



Important: CDH requires IPv4. IPv6 is not supported.



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

Procedure

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:255.255.2
48.0
...
```

- 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
```



Important: If the host command is not installed on your system, then install it by running the following command:

- RHEL:

```
yum install bind-utils
```

- Ubuntu:

```
apt install bind9-host
```

- SLES:

```
zypper in bind-utils
```

Setting SELinux Mode

Security-Enhanced Linux (SELinux) allows you to set access control through policies. If you are having trouble deploying Runtime or CDH with your policies, set SELinux in permissive mode on each host before you deploy Runtime or CDH on your cluster.

About this task



Note: CDP Private Cloud Base, with the exception of Cloudera Navigator Encrypt, is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in enforcing mode. Cloudera is not responsible for SELinux policy development, support, or enforcement. If you experience issues running Cloudera software with SELinux enabled, contact your OS provider for assistance.

If you are using SELinux in enforcing mode, Cloudera Support can request that you disable SELinux or change the mode to permissive to rule out SELinux as a factor when investigating reported issues.

Procedure

1. Check the SELinux state:

```
getenforce
```

2. If the output is either Permissive or Disabled, you can skip this task and continue to [Disabling the Firewall](#) to disable the firewall on each host in your cluster. If the output is enforcing, continue to the next step.
3. Open the `/etc/selinux/config` file (in some systems, the `/etc/sysconfig/selinux` file).
4. Change the line `SELINUX=enforcing` to `SELINUX=permissive`.
5. Save and close the file.
6. Restart your system or run the following command to disable SELinux immediately:

```
setenforce 0
```

After you have installed and deployed Runtime or CDH, you can re-enable SELinux by changing `SELINUX=permissive` back to `SELINUX=enforcing` in `/etc/selinux/config` (or `/etc/sysconfig/selinux`), and then running the following command to immediately switch to enforcing mode:

```
setenforce 1
```

If you are having trouble getting Cloudera Software working with SELinux, contact your OS vendor for support. Cloudera is not responsible for developing or supporting SELinux policies.

Disabling the Firewall

To disable the firewall on each host in your cluster, perform the following steps on each host.

Procedure

1. If the iptables command is not installed on your system, then install it by running the following command:

- RHEL:

```
sudo yum install iptables
```

- SLES:

```
sudo zypper install iptables
```

- Ubuntu:

```
sudo apt-get install iptables
```

2. For iptables, save the existing rule set:

```
sudo iptables-save > ~/firewall.rules
```

3. Disable the firewall:

- RHEL 7 compatible:

```
sudo systemctl disable firewalld  
sudo systemctl stop firewalld
```

Enable an NTP Service

Runtime requires that you configure a Network Time Protocol (NTP) service on each machine in your cluster. Most operating systems include the ntpd service for time synchronization.

About this task

RHEL 7 compatible operating systems use chronyd by default instead of ntpd. If chronyd is running (on any OS), Cloudera Manager uses it to determine whether the host clock is synchronized. Otherwise, Cloudera Manager uses ntpd.



Note: If you are using ntpd to synchronize your host clocks, but chronyd is also running, Cloudera Manager relies on chronyd to verify time synchronization, even if it is not synchronizing properly. This can result in Cloudera Manager reporting clock offset errors, even though the time is correct.

To fix this, either configure and use chronyd or disable it and remove it from the hosts.

To use ntpd for time synchronization:

Before you begin

Procedure

1. Install the ntp package:

- RHEL compatible:

```
yum install ntp
```

2. Edit the `/etc/ntp.conf` file to add NTP servers, as in the following example:

```
server 0.pool.ntp.org
server 1.pool.ntp.org
server 2.pool.ntp.org
```

3. Start the `ntpd` service:

- RHEL 7 Compatible:

```
sudo systemctl start ntpd
```

4. Configure the `ntpd` service to run at boot:

- RHEL 7 Compatible:

```
sudo systemctl enable ntpd
```

5. Synchronize the system clock to the NTP server:

```
ntpdate -u <ntp_server>
```

6. Synchronize the hardware clock to the system clock:

```
hwclock --systohc
```

Impala Requirements

To perform as expected, Impala depends on the availability of the software, hardware, and configurations described in the following sections.

Product Compatibility Matrix

The ultimate source of truth about compatibility between various versions of Cloudera Runtime, Cloudera Manager, and various Runtime components is the Product Compatibility Matrix.

Supported Operating Systems

The relevant supported operating systems and versions for Impala are the same as for the corresponding Cloudera Runtime platforms. For details, see the *Operating System Requirements* topic.

Hive Metastore and Related Configuration

Impala can interoperate with data stored in Hive, and uses the same infrastructure as Hive for tracking metadata about schema objects such as tables and columns. The following components are prerequisites for Impala:

To install the metastore:

1. Install a MySQL or PostgreSQL database. Start the database if it is not started after installation.
2. Download the MySQL Connector or the PostgreSQL connector and place it in the `/usr/share/java/` directory.
3. Use the appropriate command line tool for your database to create the metastore database.
4. Use the appropriate command line tool for your database to grant privileges for the metastore database to the hive user.
5. Modify `hive-site.xml` to include information matching your particular database: its URL, username, and password. You will copy the `hive-site.xml` file to the Impala Configuration Directory later in the Impala installation process.

Java Dependencies

Although Impala is primarily written in C++, it does use Java to communicate with various Hadoop components:

- The officially supported JVMs for Impala are the OpenJDK JVM and Oracle JVM. Other JVMs might cause issues, typically resulting in a failure at `impalad` startup. In particular, the JamVM used by default on certain levels of Ubuntu systems can cause `impalad` to fail to start.
- Internally, the `impalad` daemon relies on the `JAVA_HOME` environment variable to locate the system Java libraries. Make sure the `impalad` service is not run from an environment with an incorrect setting for this variable.
- All Java dependencies are packaged in the `impala-dependencies.jar` file, which is located at `/usr/lib/impala/lib/`. These map to everything that is built under `fe/target/dependency`.

Networking Configuration Requirements

As part of ensuring best performance, Impala attempts to complete tasks on local data, as opposed to using network connections to work with remote data. To support this goal, Impala matches the hostname provided to each Impala daemon with the IP address of each DataNode by resolving the hostname flag to an IP address. For Impala to work with local data, use a single IP interface for the DataNode and the Impala daemon on each machine. Ensure that the Impala daemon's hostname flag resolves to the IP address of the DataNode. For single-homed machines, this is usually automatic, but for multi-homed machines, ensure that the Impala daemon's hostname resolves to the correct interface. Impala tries to detect the correct hostname at start-up, and prints the derived hostname at the start of the log in a message of the form:

```
Using hostname: impala-daemon-1.example.com
```

In the majority of cases, this automatic detection works correctly. If you need to explicitly set the hostname, do so by setting the `--hostname` flag.

Hardware Requirements

The memory allocation should be consistent across Impala executor nodes. A single Impala executor with a lower memory limit than the rest can easily become a bottleneck and lead to suboptimal performance.

This guideline does not apply to coordinator-only nodes.

Hardware Requirements for Optimal Join Performance

During join operations, portions of data from each joined table are loaded into memory. Data sets can be very large, so ensure your hardware has sufficient memory to accommodate the joins you anticipate completing.

While requirements vary according to data set size, the following is generally recommended:

- CPU

Impala version 2.2 and higher uses the SSE3 instruction set, which is included in newer processors.



Note: This required level of processor is the same as in Impala version 1.x. The Impala 2.0 and 2.1 releases had a stricter requirement for the SSE4.1 instruction set, which has now been relaxed.

- Memory

128 GB or more recommended, ideally 256 GB or more. If the intermediate results during query processing on a particular node exceed the amount of memory available to Impala on that node, the query writes temporary work data to disk, which can lead to long query times. Note that because the work is parallelized, and intermediate results for aggregate queries are typically smaller than the original data, Impala can query and join tables that are much larger than the memory available on an individual node.

- JVM Heap Size for Catalog Server

4 GB or more recommended, ideally 8 GB or more, to accommodate the maximum numbers of tables, partitions, and data files you are planning to use with Impala.

- Storage

DataNodes with 12 or more disks each. I/O speeds are often the limiting factor for disk performance with Impala. Ensure that you have sufficient disk space to store the data Impala will be querying.

User Account Requirements

For user account requirements, see the topic User Account Requirements in the Impala documentation.

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)
- sed
- systemctl
- /sbin/chkconfig (RHEL/CentOS/Oracle)
- id
- rm
- mv
- chown
- install

Ports

Cloudera Manager, Cloudera Runtime components, managed services, and third-party components use the ports listed in the tables that follow.

Before you deploy Cloudera Manager, Cloudera Runtime, 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.

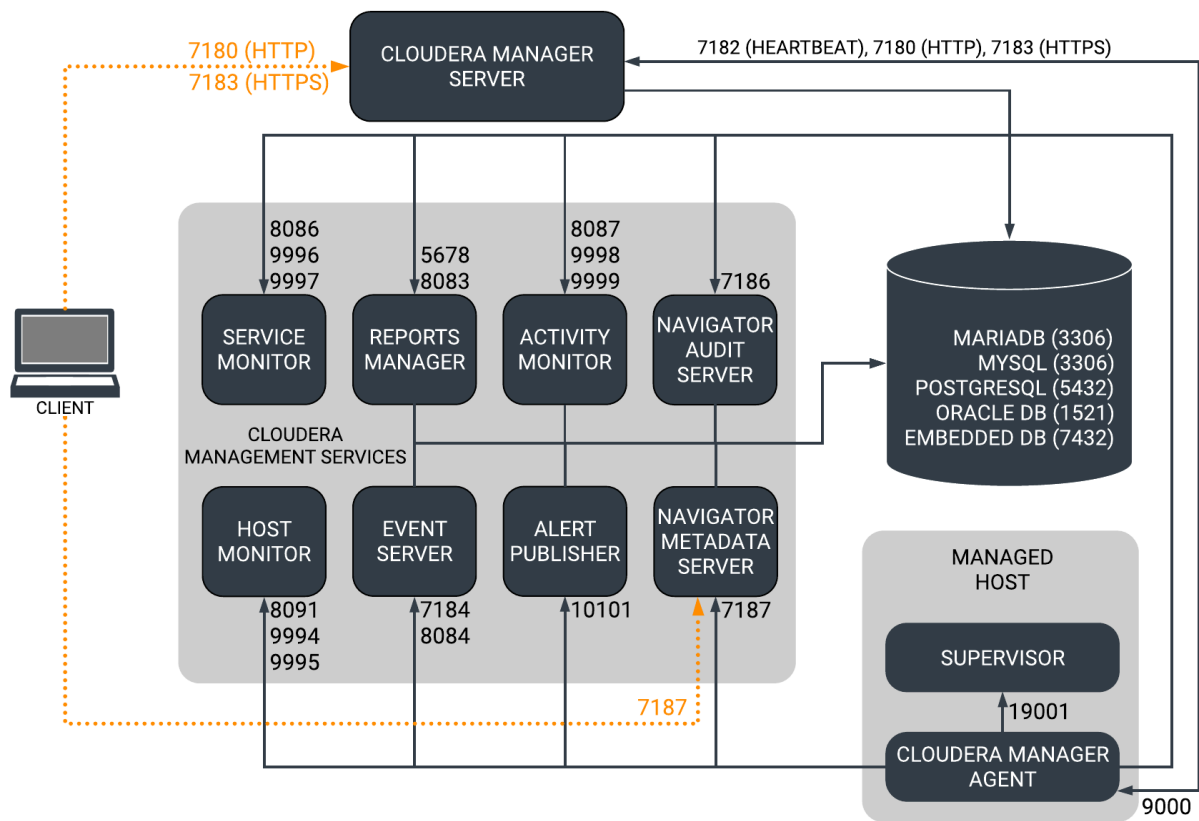
In the tables in the subsections that follow, the Access Requirement column for each port is usually either "Internal" or "External." In this context, "Internal" means that the port is used only for communication among the components (for example the JournalNode ports in an HA configuration); "External" 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

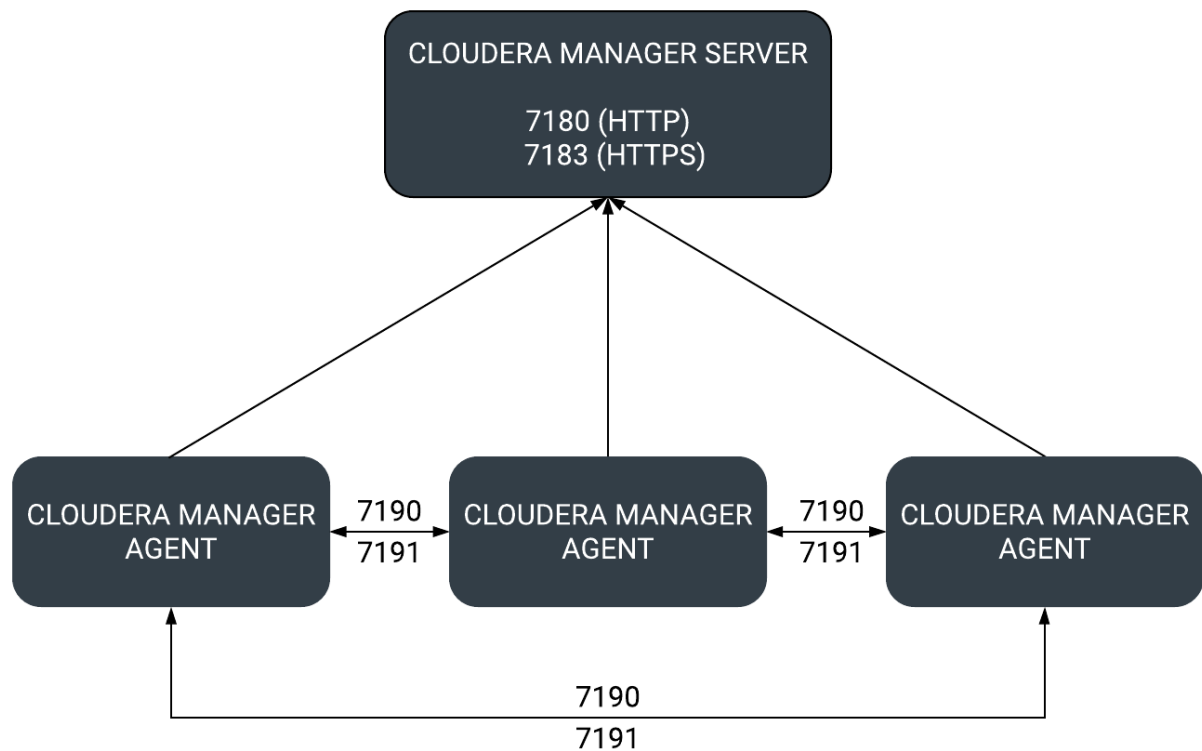
The diagrams and tables below provide an overview of some of the ports used by Cloudera Manager and Cloudera Management Service roles.

Figure 1: Ports Used by Cloudera Manager



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 14: External Ports

Component	Service	Port	Configuration	Description
Cloudera Manager Server	HTTP (Web UI)	7180	AdministrationSettingsCategory and AddressesHTTP Port for Admin Console	HTTP Port used by the web console.
	HTTPS (Web UI)	7183	AdministrationSettingsCategory and AddressesHTTPS Port for Admin Console	HTTPS 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 Manager Agent	HTTP (Debug)	9000	/etc/cloudera-scm-agent/config.ini	
Backup and Disaster Recovery	HTTP (Web UI)	7180	AdministrationSettingsCategory and AddressesHTTP Port for Admin Console	HTTP Port for communication to peer (source) Cloudera Manager.
	HTTPS (Web UI)	7183	AdministrationSettingsCategory and AddressesHTTPS Port for Admin Console	HTTPS Port for communication to peer (source) Cloudera Manager when HTTPS is enabled.

Component	Service	Port	Configuration	Description
	HDFS NameNode	8020	HDFS serviceConfigurationCategoryPorts and AddressesNameNode 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 serviceConfigurationCategoryPorts and AddressesDataNode 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	ClustersCloudera Management ServiceCategoryPorts and AddressesTelemetry Publisher Server Port	The port where the Telemetry Publisher Server listens for requests
Telemetry Publisher	HTTP (Debug)	10111	ClustersCloudera Management ServiceCategoryPorts and AddressesTelemetry Publisher Web UI Port	The port where Telemetry Publisher starts a debug web server. Set to -1 to disable debug server.

Table 15: Internal Ports

Component	Service	Port	Configuration	Description
Cloudera Manager Server	Avro (RPC)	7182	AdministrationSettingsCategoryPorts and AddressesAgent 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	HostsAll HostsConfigurationP2P Parcel Distribution Port	Used to distribute parcels to cluster hosts during installation and upgrade operations.
Cloudera Manager Agent	HTTP (Debug)	The value set for the list ening_port parameter in the /etc/clo udera-scm-ag ent/config.ini file, plus 1.	Not directly configurable. For example, the default external port is 9000. Therefore the default internal port is 9001.	
Event Server	Custom protocol	7184	Cloudera Management ServiceConfigurationCategoryPorts and AddressesEvent Publish Port	Port on which the Event Server listens for the publication of events.
	Custom protocol	7185	Cloudera Management ServiceConfigurationCategoryPorts and AddressesEvent Query Port	Port on which the Event Server listens for queries for events.

Component	Service	Port	Configuration	Description
	HTTP (Debug)	8084	Cloudera Management ServiceConfigurationCategoryPorts and AddressesEvent 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 ServiceConfigurationCategoryPorts and AddressesAlerts: Listen Port	Port where the Alert Publisher listens for internal API requests.
Service Monitor	HTTP (Debug)	8086	Cloudera Management ServiceConfigurationCategoryPorts and AddressesService Monitor Web UI Port	Port for Service Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)		Cloudera Management ServiceConfigurationCategoryPorts and AddressesService Monitor Web UI HTTPS Port	Port for Service Monitor's HTTPS Debug page.
	Custom protocol	9997	Cloudera Management ServiceConfigurationCategoryPorts and AddressesService Monitor Listen Port	Port where Service Monitor is listening for agent messages.
	Internal query API (Avro)	9996	Cloudera Management ServiceConfigurationCategoryPorts and AddressesService Monitor Nozzle Port	Port where Service Monitor's query API is exposed.
Activity Monitor	HTTP (Debug)	8087	Cloudera Management ServiceConfigurationCategoryPorts and AddressesActivity Monitor Web UI Port	Port for Activity Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)		Cloudera Management ServiceConfigurationCategoryPorts and AddressesActivity Monitor Web UI HTTPS Port	Port for Activity Monitor's HTTPS Debug page.
	Custom protocol	9999	Cloudera Management ServiceConfigurationCategoryPorts and AddressesActivity Monitor Listen Port	Port where Activity Monitor is listening for agent messages.
	Internal query API (Avro)	9998	Cloudera Management ServiceConfigurationCategoryPorts and AddressesActivity Monitor Nozzle Port	Port where Activity Monitor's query API is exposed.
Host Monitor	HTTP (Debug)	8091	Cloudera Management ServiceConfigurationCategoryPorts and AddressesHost Monitor Web UI Port	Port for Host Monitor's Debug page. Set to -1 to disable the debug server.
	HTTPS (Debug)	9091	Cloudera Management ServiceConfigurationCategoryPorts and AddressesHost Monitor Web UI HTTPS Port	Port for Host Monitor's HTTPS Debug page.
	Custom protocol	9995	Cloudera Management ServiceConfigurationCategoryPorts and AddressesHost Monitor Listen Port	Port where Host Monitor is listening for agent messages.
	Internal query API (Avro)	9994	Cloudera Management ServiceConfigurationCategoryPorts and AddressesHost Monitor Nozzle Port	Port where Host Monitor's query API is exposed.

Component	Service	Port	Configuration	Description
Reports Manager	Queries (Thrift)	5678	Cloudera Management ServiceConfigurationCategoryPorts and AddressesReports Manager Server Port	The port where Reports Manager listens for requests.
	HTTP (Debug)	8083	Cloudera Management ServiceConfigurationCategoryPorts and AddressesReports Manager Web UI Port	The port where Reports Manager starts a debug web server. Set to -1 to disable debug server.
HTTP (Debug)	8089	Cloudera Management ServiceConfigurationCategoryPorts and AddressesNavigator 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 Key Trustee Server

The Cloudera Navigator Key Trustee Server uses certain ports to store and retrieve encryption information and information required for high availability.

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 serviceConfigurationCategoryPorts and AddressesKey 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 serviceConfigurationCategoryPorts and AddressesKey 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 Cloudera Runtime Components

Cloudera Runtime components use a number of ports for associated services.

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 16: External Ports

Component	Service	Port	Configuration	Comment
Apache Atlas	Non-SSL	31000	atlas.server.http.port	
	SSL	31443	atlas.server.https.port	This port is used only when Atlas is in SSL mode.
Apache Hadoop HDFS	DataNode	9866	dfs.datanode.address	DataNode server address and port for data transfer
		9864	dfs.datanode.http.address	DataNode HTTP server port

Component	Service	Port	Configuration	Comment
	NameNode	9865	dfs.datanode.https.address	DataNode HTTPS server port
		9867	dfs.datanode.ipc.address	DataNode IPC server port
		8020	fs.default.name or fs.defaultFS	fs.default.name is deprecated (but still works)
		8022	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.
		9870	dfs.http.address or dfs.namenode.http-address	dfs.http.address is deprecated (but still works)
		9871	dfs.https.address or dfs.namenode.https-address	dfs.https.address is deprecated (but still works)
	NFS gateway	2049		nfs port (nfs3.server.port)
		4242		mountd port (nfs3.mountd.port)
		111		portmapper rpcbind port
		50079	nfs.http.port	The NFS gateway daemon uses this port to serve metrics. The port is configurable on versions 5.10 and higher.
		50579	nfs.https.port	The NFS gateway daemon uses this port to serve metrics. The port is configurable on versions 5.10 and higher.
	HttpFS	14000		HttpFS server port
		14001		HttpFS admin port
Apache Hadoop YARN (MRv2)	ResourceManager	8032	yarn.resourcemanager.address	
		8033	yarn.resourcemanager.admin.address	
		8088	yarn.resourcemanager.webapp.address	
		8090	yarn.resourcemanager.webapp.https.address	
	NodeManager	8042	yarn.nodemanager.webapp.address	
		8044	yarn.nodemanager.webapp.https.address	
	JobHistory Server	19888	mapreduce.jobhistory.webapp.address	
		19890	mapreduce.jobhistory.webapp.https.address	
	ApplicationMaster			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.
Apache Flume	Flume Agent	41414		

Component	Service	Port	Configuration	Comment
Apache Hadoop KMS	Key Management Server	16000	kms_http_port	Applies to both Java KeyStore KMS and Key Trustee KMS.
Apache HBase	Master	16000	hbase.master.port	IPC
		16010	hbase.master.info.port	HTTP
	RegionServer	16020	hbase.regionserver.port	IPC
		16030	hbase.regionserver.info.port	HTTP
	REST	20550	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		
	Thrift Server	9090	Pass -p <port> on CLI	
	Thrift Server	9095		
		9090	Pass --port <port> on CLI	
	Lily HBase Indexer	11060		
Apache Hive	Metastore	9083		
	HiveServer2	10000	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.
	HiveServer2 Web User Interface (UI)	10002	hive.server2.webui.port in hive-site.xml	
Hue	Server	8888		
	Load Balancer	8889		
Apache Impala	Impala Daemon	21000		Used to transmit commands and receive results by impala-shell and version 1.2 of the Cloudera ODBC driver.
		21050		Used to transmit commands and receive results by applications, such as Business Intelligence tools, using JDBC, the Beeswax query editor in Hue, and version 2.0 or higher of the Cloudera ODBC driver.
		25000		Impala web interface for administrators to monitor and troubleshoot.
	StateStore Daemon	25010		StateStore web interface for administrators to monitor and troubleshoot.
	Catalog Daemon	25020		Catalog service web interface for administrators to monitor and troubleshoot.
Apache Kafka	Kafka Broker	9092	port	The primary communication port used by producers and consumers; also used for inter-broker communication.

Component	Service	Port	Configuration	Comment
		9093	ssl_port	A secured communication port used by producers and consumers; also used for inter-broker communication.
	Kafka Connect	38083	rest.port	Kafka Connect Rest Port
		38085	secure.rest.port	Kafka Connect Secure Rest Port
Apache Kudu	Master	7051		Kudu Master RPC port
		8051		Kudu Master HTTP server port
	TabletServer	7050		Kudu TabletServer RPC port
		8050		Kudu TabletServer HTTP server port
Apache Oozie	Oozie Server	11000	OOZIE_HTTP_PORT in oozie-env.sh	HTTP
		11443		HTTPS
Apache Ranger	Non-SSL	6080	ranger.service.http.port	
	SSL	6182	ranger.service.https.port	This port is used only when Ranger is in SSL mode.
	Admin Unix Auth Service Port	5151	ranger.unixauth.service.port	
Apache Solr	Solr Server	8983		HTTP port for all Solr-specific actions, update/query.
	Solr Server	8985		HTTPS port for all Solr-specific actions, update/query.
Apache Spark	Shuffle service	7377	spark.shuffle.service.port	
	History Server	18081	spark.history.ui.port	
	History Server with TLS	18488	spark.ssl.historyServer.port	
Apache Sqoop	Metastore	16000	sqoop.metastore.server.port	
Apache ZooKeeper	Server (with CDH or Cloudera Manager)	2181	clientPort	Client port
Cruise Control	Cruise Control Server	8899	webserver.http.port	This is the main port that enables access to the Cruise Control Server
Livy	Livy Server Web UI	8998	livy.server.port	
	Livy Thrift Server	10090	livy.server.thrift.port	
Schema Registry	Schema Registry Server	7788	schema.registry.port	REST endpoint for Schema Registry.
		7789	schema.registry.adminPort	Page for monitoring the Schema Registry service to determine for example the health state and CPU usage.
		7790	schema.registry.ssl.port	When SSL is enabled, REST endpoint for Schema Registry.
		7791	schema.registry.ssl.adminPort	When SSL is enabled, the page for monitoring the Schema Registry service to determine for example the health state and CPU usage.
Streams Messaging Manager	Streams Messaging Manager Rest Admin Server	8585	streams.messaging.manager.port	Streams Messaging Manager Port
		8587	streams.messaging.manager.ssl.port	Streams Messaging Manager Port (SSL)

Component	Service	Port	Configuration	Comment
		8586	streams.messaging.manager.adminPort	Streams Messaging Manager Admin Port
		8588	streams.messaging.manager.ssl.adminPort	Streams Messaging Manager Admin Port (SSL)
	Streams Messaging Manager UI Server	9991	streams.messaging.manager.ui.port	The port on which server accepts connections. This port is used for both secured and unsecured connections.
Streams Replication Manager	SRM Service	6670	streams.replication.manager.service.port	SRM Service port.
		6671	streams.replication.manager.service.ssl.port	SRM Service port. when SSL is enabled.

Table 17: Internal Ports

Component	Service	Port	Configuration	Comment
Apache Hadoop HDFS	Secondary NameNode	9868	dfs.secondary.http.address or dfs.namenode.secondary.http-address	dfs.secondary.http.address is deprecated (but still works)
		9869	dfs.secondary.https.address	
	JournalNode	8485	dfs.namenode.shared.edits.dir	
		8480	dfs.journalnode.http-address	
		8481	dfs.journalnode.https-address	
	Failover Controller	8019		Used for NameNode HA
Apache Hadoop YARN (MRv2)	ResourceManager	8030	yarn.resourcemanager.scheduler.address	
		8031	yarn.resourcemanager.resource-tracker.address	
	NodeManager	8040	yarn.nodemanager.localizer.address	
		8041	yarn.nodemanager.address	
	JobHistory Server	10020	mapreduce.jobhistory.address	
		10033	mapreduce.jobhistory.admin.address	
	Shuffle HTTP	13562	mapreduce.shuffle.port	
	Queue Manager	8082	queuemanager_webapp_port	
	Config Store/Service	8080	Set this configuration in the config.yml file for the service.	Reconfiguring this in a production environment is not recommended.
Apache Hadoop KMS	Key Management Server	8081	adminConnectorsPort	Set this configuration in the config.yml file for the service.
		16001	kms_admin_port	Applies to both Java KeyStore KMS and Key Trustee KMS.
Apache HBase	HQuorumPeer	2181	hbase.zookeeper.property.clientPort	HBase-managed ZooKeeper mode
		2888	hbase.zookeeper.peerport	HBase-managed ZooKeeper mode
		3888	hbase.zookeeper.leaderport	HBase-managed ZooKeeper mode
Apache Impala	Impala Daemon	22000		Internal use only. Impala daemons use this port to communicate with each other.

Component	Service	Port	Configuration	Comment
		23000		Internal use only. Impala daemons listen on this port for updates from the statestore daemon.
	StateStore Daemon	24000		Internal use only. The statestore daemon listens on this port for registration/unregistration requests.
	Catalog Daemon	23020		Internal use only. The catalog daemon listens on this port for updates from the statestore daemon.
		26000		Internal use only. The catalog service uses this port to communicate with the Impala daemons.
Apache Kafka	Kafka Broker	9092	port	The primary communication port used by producers and consumers; also used for inter-broker communication.
		9093	ssl_port	A secured communication port used by producers and consumers; also used for inter-broker communication.
		9393	jmx_port	Internal use only. Used for administration via JMX.
		9394	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.
	Kafka Connect	38084	metrics.jetty.server.port	Metrics Jetty Server Port
	Kafka MirrorMaker	24042	jmx_port	Internal use only. Used to administer the producer and consumer of the MirrorMaker.
Apache Solr	Solr Server	8993		Infra-Solr HTTP port
	Solr Server	8995		Infra-Solr HTTPS port
Apache ZooKeeper	Server (with CDH only)	2888	X in server.N =host:X:Y	Peer
	Server (with CDH only)	3888	X in server.N =host:X:Y	Peer
	Server (with CDH and Cloudera Manager)	3181	X in server.N =host:X:Y	Peer
	Server (with CDH and Cloudera Manager)	4181	X in server.N =host:X:Y	Peer

Component	Service	Port	Configuration	Comment
	ZooKeeper JMX port	9010		<p>ZooKeeper will also use another randomly selected port for RMI. To allow Cloudera Manager to monitor ZooKeeper, you must do one of the following:</p> <ul style="list-style-type: none"> • Open up all ports when the connection originates from the Cloudera Manager Server • Do the following: <ol style="list-style-type: none"> 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: <code>-Dcom.sun.management.jmxremote.rmi.port=9011</code> 4. Restart ZooKeeper.
Apache Zeppelin	Zeppelin Server	8885	zeppelin.server.port	
	Zeppelin Server (SSL)	8886	zeppelin.server.ssl.port	
Streams Messaging Manager	Streams Messaging Manager Rest Admin Server	6670	streams.replication.manager.port	Streams Replication Manager rest port
		6671	streams.replication.manager.port	Streams Replication Manager rest port on SSL
		7180	cm.metrics.port	Cloudera Manager's HTTP port.
		7183	cm.metrics.port	Cloudera Manager's HTTPS port
		9997	cm.metrics.service.monitor.port	Cloudera Manager Service Monitor port
		38083	kafka.connect.port	Kafka Connect port
		3306	streams.messaging.manager.storage.connector.port	Streams Messaging Manager database port

Ports Used by DistCp

DistCp uses various ports for HDFS and HttpFS services.

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	<code>fs.default.name</code>	<code>fs.default.name</code>
					or <code>fs.defaultFS</code>	is deprecated (but still works)
	DataNode	Secure	1004	External	<code>dfs.datanode.address</code>	
	DataNode		50010	External	<code>dfs.datanode.address</code>	
WebHDFS	NameNode		50070	External	<code>dfs.http.address</code>	<code>dfs.http.address</code>
					or <code>dfs.namenode.http-address</code>	is deprecated (but still works)
	DataNode	Secure	1006	External	<code>dfs.datanode.http.address</code>	
HttpFS	web		14000			

Ports Used by Third-Party Components

Third-party components such as PostgreSQL and LDAP use a number of ports for associated services.

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		

Component	Service	Qualifier	Port	Protocol	Access Requirement	Configuration	Comment
	ganglia-web		80	TCP	External	Via Apache <code>httpd</code>	
Kerberos	KRB5 KDC Server	Secure	88	UDP/TCP	External	<code>kdc_ports</code> and <code>kdc_tcp_ports</code> in either the <code>[kdcdefaults]</code> or <code>[realms]</code> sections of <code>kdc.conf</code>	By default only UDP
	KRB5 Admin Server	Secure	749	TCP	External	<code>kadmind_port</code> in the <code>[realms]</code> section of <code>kdc.conf</code>	
	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		

Runtime Cluster Hosts and Role Assignments

Cluster hosts can be broadly described as master hosts, utility hosts, gateway hosts, or worker hosts.

- 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 one or more Hive Metastores.
- 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 Runtime is used in a production environment.

The following tables describe the recommended role allocations for different cluster sizes. Note that these configurations take into account services dependencies that might not be obvious. For example, running Atlas or Ranger requires also running HBase, Kafka, Solr, and ZooKeeper. For details see [Service Dependencies in Cloudera Manager](#).

3 - 10 Worker Hosts without High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: <ul style="list-style-type: none"> • NameNode • YARN ResourceManager • JobHistory Server • ZooKeeper • Kudu master • Spark History Server • HBase master • Schema Registry 	One host for all Utility and Gateway roles: <ul style="list-style-type: none"> • Secondary NameNode • Cloudera Manager • Cloudera Manager Management Service • Cruise Control • Hive Metastore • HiveServer2 • Impala Catalog Server • Impala StateStore • Hue • Oozie • Gateway configuration • HBase backup master • Ranger Admin, Tagsync, Usersync servers • Atlas server • Solr server (CDP-INFRA-SOLR instance to support Atlas) • Streams Messaging Manager • Streams Replication Manager Service • ZooKeeper 		3 - 10 Worker Hosts: <ul style="list-style-type: none"> • DataNode • NodeManager • Impalad • Kudu tablet server • Kafka Broker • Kafka Connect • HBase RegionServer • Solr server (For Cloudera Search) • Streams Replication Manager Driver • ZooKeeper (Recommend 3 servers total)

3 - 20 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
<p>Master Host 1:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper JobHistory Server Kudu master HBase master Schema Registry <p>Master Host 2:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper Kudu master HBase master Schema Registry <p>Master Host 3:</p> <ul style="list-style-type: none"> Kudu master (Kudu requires an odd number of masters for HA.) Spark History Server JournalNode (requires dedicated disk) ZooKeeper 	<p>Utility Host 1:</p> <ul style="list-style-type: none"> Cloudera Manager Cloudera Manager Management Service Cruise Control Hive Metastore Impala Catalog Server Impala StateStore Oozie Ranger Admin, Tagsync, Usersync servers Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) Streams Messaging Manager Streams Replication Manager Service <p>Utility Host 2:</p> <ul style="list-style-type: none"> Hive Metastore Ranger Admin server Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) 	<p>One or more Gateway Hosts:</p> <ul style="list-style-type: none"> Hue HiveServer2 Gateway configuration 	<p>3 - 20 Worker Hosts:</p> <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server Kafka Broker (Recommend 3 brokers minimum) Kafka Connect HBase RegionServer Solr server (For Cloudera Search, recommend 3 servers minimum) Streams Replication Manager Driver

20 - 80 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
<p>Master Host 1:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper Kudu master HBase master Schema Registry <p>Master Host 2:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper Kudu master HBase master Schema Registry <p>Master Host 3:</p> <ul style="list-style-type: none"> ZooKeeper JournalNode JobHistory Server Spark History Server Kudu master HBase master 	<p>Utility Host 1:</p> <ul style="list-style-type: none"> Cloudera Manager Cruise Control Hive Metastore Ranger Admin server Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) Streams Messaging Manager Streams Replication Manager Service <p>Utility Host 2:</p> <ul style="list-style-type: none"> Cloudera Manager Management Service Hive Metastore Impala Catalog Server Impala StateStore Oozie Ranger Admin, Tagsync, Usersync servers Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) 	<p>One or more Gateway Hosts:</p> <ul style="list-style-type: none"> Hue HiveServer2 Gateway configuration 	<p>20 - 80 Worker Hosts:</p> <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server Kafka Broker (Recommend 3 brokers minimum) Kafka Connect HBase RegionServer Solr server (For Cloudera Search, recommend 3 servers minimum) Streams Replication Manager Driver

80 - 200 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
<p>Master Host 1:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper Kudu master HBase master Schema Registry <p>Master Host 2:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController YARN ResourceManager ZooKeeper Kudu master HBase master Schema Registry <p>Master Host 3:</p> <ul style="list-style-type: none"> ZooKeeper JournalNode JobHistory Server Spark History Server Kudu master HBase master 	<p>Utility Host 1:</p> <ul style="list-style-type: none"> Cloudera Manager Cruise Control Streams Messaging Manager Streams Replication Manager Service <p>Utility Host 2:</p> <ul style="list-style-type: none"> Hive Metastore Impala Catalog Server Impala StateStore Oozie <p>Utility Host 3:</p> <ul style="list-style-type: none"> Host Monitor <p>Utility Host 4:</p> <ul style="list-style-type: none"> Ranger Admin, Tagsync, Usersync servers Atlas server Solr server <p>Utility Host 5:</p> <ul style="list-style-type: none"> Hive Metastore Ranger Admin server Atlas server Solr server <p>Utility Host 6:</p> <ul style="list-style-type: none"> Reports Manager <p>Utility Host 7:</p> <ul style="list-style-type: none"> Service Monitor 	<p>One or more Gateway Hosts:</p> <ul style="list-style-type: none"> Hue HiveServer2 Gateway configuration 	<p>80 - 200 Worker Hosts:</p> <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server (Recommend 100 tablet servers maximum) Kafka Broker (Recommend 3 brokers minimum) Kafka Connect HBase RegionServer Solr server (For Cloudera Search, recommend 3 servers minimum) Streams Replication Manager Driver

200 - 500 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
<p>Master Host 1:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController ZooKeeper Kudu master HBase master <p>Master Host 2:</p> <ul style="list-style-type: none"> NameNode JournalNode FailoverController ZooKeeper Kudu master HBase master <p>Master Host 3:</p> <ul style="list-style-type: none"> YARN ResourceManager ZooKeeper JournalNode Kudu master HBase master Schema Registry <p>Master Host 4:</p> <ul style="list-style-type: none"> YARN ResourceManager ZooKeeper JournalNode Schema Registry <p>Master Host 5:</p> <ul style="list-style-type: none"> JobHistory Server Spark History Server ZooKeeper JournalNode <p>We recommend no more than three masters for Kudu and HBase.</p>	<p>Utility Host 1:</p> <ul style="list-style-type: none"> Cloudera Manager Cruise Control Streams Messaging Manager Streams Replication Manager Service <p>Utility Host 2:</p> <ul style="list-style-type: none"> Hive Metastore Impala Catalog Server Impala StateStore Oozie <p>Utility Host 3:</p> <ul style="list-style-type: none"> Host Monitor <p>Utility Host 4:</p> <ul style="list-style-type: none"> Ranger Admin, Tagsync, Usersync servers Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) <p>Utility Host 5:</p> <ul style="list-style-type: none"> Hive Metastore Ranger Admin server Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) <p>Utility Host6:</p> <ul style="list-style-type: none"> Reports Manager <p>Utility Host 7:</p> <ul style="list-style-type: none"> Service Monitor 	<p>One or more Gateway Hosts:</p> <ul style="list-style-type: none"> Hue HiveServer2 Gateway configuration 	<p>200 - 500 Worker Hosts:</p> <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server (Recommend 100 tablet servers maximum) Kafka Broker (Recommend 3 brokers minimum) Kafka Connect HBase RegionServer Solr server (For Cloudera Search, recommend 3 servers minimum) Streams Replication Manager Driver

500 -1000 Worker Hosts with High Availability

Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: <ul style="list-style-type: none"> NameNode JournalNode FailoverController ZooKeeper Kudu master HBase master Master Host 2: <ul style="list-style-type: none"> NameNode JournalNode FailoverController ZooKeeper Kudu master HBase master Master Host 3: <ul style="list-style-type: none"> YARN ResourceManager ZooKeeper JournalNode Kudu master HBase master Schema Registry Master Host 4: <ul style="list-style-type: none"> YARN ResourceManager ZooKeeper JournalNode Schema Registry Master Host 5: <ul style="list-style-type: none"> JobHistory Server Spark History Server ZooKeeper JournalNode <p>We recommend no more than three masters for Kudu and HBase.</p>	Utility Host 1: <ul style="list-style-type: none"> Cloudera Manager Cruise Control Streams Messaging Manager Streams Replication Manager Service Utility Host 2: <ul style="list-style-type: none"> Hive Metastore Impala Catalog Server Impala StateStore Oozie Utility Host 3: <ul style="list-style-type: none"> Host Monitor Utility Host 4: <ul style="list-style-type: none"> Ranger Admin, Tagsync, Usersync servers Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) Utility Host 5: <ul style="list-style-type: none"> Hive Metastore Ranger Admin server Atlas server Solr server (CDP-INFRA-SOLR instance to support Atlas) Utility Host 6: <ul style="list-style-type: none"> Reports Manager Utility Host 7: <ul style="list-style-type: none"> Service Monitor 	One or more Gateway Hosts: <ul style="list-style-type: none"> Hue HiveServer2 Gateway configuration 	500 - 1000 Worker Hosts: <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server (Recommend 100 tablet servers maximum) Kafka Broker (Recommend 3 brokers minimum) Kafka Connect HBase RegionServer Solr server (For Cloudera Search, recommend 3 servers minimum) Streams Replication Manager Driver

Related Information

[Configuring HMS for high availability](#)

Allocating Hosts for Key Trustee Server and Key Trustee KMS

If you are enabling data-at-rest encryption for a Cloudera Runtime 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 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.

Service Dependencies in Cloudera Manager

The following tables list service dependencies that exist between various services in a Cloudera Manager deployment. As you configure services for Cloudera Manager, refer to the tables below for the appropriate version.

Service dependencies for Spark 2 on YARN and Cloudera Data Science Workbench are listed separately.

Table 18: Version 7.0 Service Dependencies

Service	Dependencies	Optional Dependencies
ADLS Connector		
Atlas	<ul style="list-style-type: none"> HDFS HBase Kafka (Kafka broker role only) Solr 	Ranger
Data Context Connector		
HBase	<ul style="list-style-type: none"> HDFS ZooKeeper 	<ul style="list-style-type: none"> Atlas Ranger
HDFS		<ul style="list-style-type: none"> ADLS Connector or S3 Connector KMS, Thales KMS, Key Trustee, or Luna KMS Ranger ZooKeeper
Hive	HDFS	<ul style="list-style-type: none"> Atlas HBase Kudu Ranger Spark on YARN YARN ZooKeeper
Hive-on-Tez	<ul style="list-style-type: none"> HDFS Hive Tez 	<ul style="list-style-type: none"> Atlas HBase Ranger YARN ZooKeeper
Hue	<ul style="list-style-type: none"> HDFS Hive 	<ul style="list-style-type: none"> Atlas HBase Hive-on-Tez Impala Oozie Solr ZooKeeper
Impala	<ul style="list-style-type: none"> HDFS Hive 	<ul style="list-style-type: none"> Atlas HBase Kudu Ranger YARN ZooKeeper
Kafka	ZooKeeper	<ul style="list-style-type: none"> HDFS Ranger

Service	Dependencies	Optional Dependencies
Key-Value Store Indexer	<ul style="list-style-type: none"> HBase Solr 	Ranger
Kudu		Ranger
Livy	<ul style="list-style-type: none"> Spark-on-YARN YARN 	Hive
Oozie	YARN	<ul style="list-style-type: none"> Hive Spark on YARN ZooKeeper
Ozone		<ul style="list-style-type: none"> HDFS Ranger
Ranger	<ul style="list-style-type: none"> HDFS Solr 	
S3 Connector		
Solr	<ul style="list-style-type: none"> HDFS ZooKeeper 	Ranger
Spark on YARN	YARN	<ul style="list-style-type: none"> Atlas HBase
Tez	YARN	
YARN	<ul style="list-style-type: none"> HDFS ZooKeeper 	Ranger
Zeppelin	<ul style="list-style-type: none"> HDFS Spark-on-YARN YARN 	<ul style="list-style-type: none"> Livy
ZooKeeper		

Custom Installation Solutions

Some installations may require custom solutions such as creating virtual images of cluster hosts, configuring a custom Java home location, or creating a Runtime cluster using a template.

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.

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 Runtime 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

Complete the steps below to create a Cloudera Manager virtual machine image.

Procedure

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 Runtime 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:

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

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

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

Complete the following steps to create a new Cloudera Manager instance from a virtual machine image.

Procedure

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:

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

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

Creating a Pre-Deployed Worker Host

Complete the steps below to create a pre-deployed worker host.

Procedure

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 7 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:

```
shasum 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:

```
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 parcelfile -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-6.0.0-1.cdh6.0.0.p0.309038 to /opt/cloudera/parcels/CDH, use the following command:

```
ln -s /opt/cloudera/parcels/CDH-6.0.0-1.cdh6.0.0.p0.309038 /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/parcelname
```

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

Instantiating a Worker Host

Complete the steps below to instantiate a worker host.

Procedure

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.
4. Start the agent service:

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

Configuring a Custom Java Home Location

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*.

About this task

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 Runtime cluster processes only. It does not affect the JDK used by other non-Cloudera processes, or gateway roles. To modify the Cloudera Manager configuration to ensure the JDK can be found:

Procedure

1. Open the Cloudera Manager Admin Console.

2. In the left-side navigation bar, click **Hosts** **Hosts Configuration**. If you are configuring the JDK location on a specific host only, click **Hosts** **All Hosts**, select the specific host that you want to configure, and click the **Configuration** tab.
3. Select **Category** **Advanced**.
4. Set the **Java Home Directory** property to the custom location.
5. Click **Save Changes**.
6. Restart all services.

Creating a Runtime 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.

About this task

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 and refer to the following tasks to create a template and a new cluster:

Procedure

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.
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.
3. Create any local repositories required for the cluster.
See Step 1: Configure a Repository for Cloudera Manager.
4. Complete the **instantiator** section of the cluster configuration JSON document to create a template.
5. Import the cluster template to the new cluster.

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.

Procedure

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" >
path_to_file/file_name.json?exportAutoConfig=true
```



Note: The cluster from which you export the JSON configuration file may have been installed using either parcels or packages. You can determine the type of installation from the JSON configuration file by examining the first section. If there are entries for `"repositories": []`, or `"products": []`, then the cluster was installed using parcels. Clusters installed using packages do not have these entries.

For example:

```
{
  "cdhVersion" : "5.15.0",
  "displayName" : "Cluster 1",
  "cmVersion" : "5.15.0",
  "repositories" : [ "http://my_cluster.com/cdh5/5.x/parcels", "http://my_cluster.com/cdh5/parcels/ ..."
  "products" : [ {
    "version" : "5.15.0-1.cdh5.15.0.p0.727055",
    "product" : "CDH"
  } ],
```

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 is installed and running.
- All required databases for Runtime services are installed. See *Step 4: Install and Configure Databases*.
- 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.

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.

About this task

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_1",
      "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"
    }
  ]
}

```

Procedure

1. To modify the template, 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-REGIONSE
RVER-BASE", "HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IM
PALAD-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-REGIONSE
RVER-BASE", "HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1-IMP
ALAD-BASE", "KAFKA-1-KAFKA_BROKER-BASE", "KUDU-1-KUDU_TSERVER-BASE", "MA
PREDUCE-1-TASKTRACKER-BASE", "SPARK_ON_YARN-1-GATEWAY-BASE", "SQOOP_CLIE
NT-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:

- Locate the entry in the hosts section of the instantiator where you want the roles to be installed.
- Copy the value of the refName to the value for hostTemplateRefName.
- Enter the hostname in the new cluster as the value for hostName. Some host sections might instead use host NameRange 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-GATEW
AY-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-BAS
E", "HIVE-1-HIVEMETASTORE-BASE", "HIVE-1-HIVESERVER2-BASE", "HUE-1-HUE_L
OAD_BALANCER-BASE", "HUE-1-HUE_SERVER-BASE", "IMPALA-1-CATALOGSERVER-BAS
E", "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-S
PARK_YARN_HISTORY_SERVER-BASE", "SQOOP-1-SQOOP_SERVER-BASE", "SQOOP_CLIE
NT-1-GATEWAY-BASE", "YARN-1-GATEWAY-BASE", "YARN-1-JOBHISTORY-BASE", "YA
RN-1-RESOURCEMANAGER-BASE", "ZOOKEEPER-1-SERVER-BASE" ]
}, {
  "refName" : "HostTemplate-1-from-myCluster-4.myDomain.com",
  "cardinality" : 1,
  "roleConfigGroupsRefNames" : [ "FLUME-1-AGENT-BASE", "HBASE-1-REGI
ONSERVER-BASE", "HDFS-1-DATANODE-BASE", "HIVE-1-GATEWAY-BASE", "IMPALA-1
-IMPALAD-BASE", "KUDU-1-KUDU_TSERVER-BASE", "MAPREDUCE-1-TASKTRACKER-BAS
E", "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-REGIO
NSERVER-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_C
LIENT-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 Runtime 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

Complete the steps below to import the cluster template.

Procedure

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
ht
tp://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 ht
tp://admin:admin@myNewCluster-1.mydomain.com:7182/api/v12/cm/importClust
erTemplate
```

```
curl -X POST -H "Content-Type: application/json" -d @myTemplate.json ht
tp://admin:admin@myNewCluster-1.mydomain.com:7182/api/v12/cm/importClust
erTemplate?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", v
ersion=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)
```

Local Package and Parcel Repositories

Cloudera hosts two types of software repositories that you can use to install products such as Cloudera Manager or Cloudera Runtime—parcel repositories and package repositories. These repositories are effective solutions in most cases, but custom installation solutions are sometimes required.

For example, 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 Runtime cluster, all hosts must run the same Runtime 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.

Understanding Package Management

Before you configure a custom package management solution in your environment, understand the concepts of package management tools and package repositories.

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), 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.

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 *Configuring a Local Package Repository*.

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`

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. 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`

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

repo id	repo name	st
atus		
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		

Configuring a Local Package Repository

You can create a package repository for Cloudera Manager either by hosting an internal web repository or by manually copying the repository files to the Cloudera Manager Server host for distribution to Cloudera Manager Agent hosts.

Creating a Permanent Internal Repository

The following sections describe how to create a permanent internal repository using Apache HTTP Server.

Setting Up a Web Server

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.

About this task

The examples in this section use Apache HTTP Server as the Web server. If you already have a Web server in your organization, you can skip to *Downloading and Publishing the Package Repository*.

Procedure

1. Install Apache HTTP Server:

RHEL / CentOS

```
sudo yum install httpd
```

2. Start Apache HTTP Server:

RHEL 7

```
sudo systemctl start httpd
```

Downloading and Publishing the Package Repository

Download the package repository for the product you want to install.

Procedure

1. Download the package repository for the product you want to install:

Cloudera Manager 7

To download the files for a Cloudera Manager release, download the repository tarball for your operating system. Then unpack the tarball, move the files to the web server directory, and modify file permissions. For example:

```
sudo mkdir -p /var/www/html/cloudera-repos/cm7
```

```
wget https://[username]:[password]@archive.cloudera.com/p/cm7/7.0  
.3/repo-as-tarball/cm7.0.3-redhat7.tar.gz
```

```
tar xvfz cm7.0.3-redhat7.tar.gz -C /var/www/html/cloudera-repos/  
cm7 --strip-components=1
```

```
sudo chmod -R ugo+rX /var/www/html/cloudera-repos/cm7
```

2. Visit the Repository URL `http://<web_server>/cloudera-repos/` in your browser and verify the files you downloaded are present. 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 packages on a one-time basis. Cloudera recommends using the same host that runs Cloudera Manager, or a gateway host.

About this task

This example uses Python SimpleHTTPServer as the Web server to host the `/var/www/html` directory, but you can use a different directory.

Procedure

1. Download the repository you need following the instructions in *Downloading and Publishing the Package Repository*.
2. Determine a port that your system is not listening on. This example uses port 8900.
3. Start a Python SimpleHTTPServer in the /var/www/html directory:

```
cd /var/www/html
python -m SimpleHTTPServer 8900
```

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

4. Visit the Repository URL `http://<web_server>:8900/cloudera-repos/` in your browser and verify the files you downloaded are present.

Configuring Hosts to Use the Internal Repository

After you establish the repository, modify the client configuration to use it.

OS	Procedure
RHEL compatible	<p>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:</p> <pre>[cloudera-repo] name=cloudera-repo baseurl=http://<web_server>/cloudera-repos/cm7 enabled=1 gpgcheck=0</pre>

Manually Install Cloudera Software Packages

This topic shows how to manually install Cloudera Manager packages. Package installations of Cloudera Runtime are not supported in CDP Private Cloud Base .

Install Cloudera Manager Packages

Cloudera Manager is installed on the Cloudera Manager Server host using packages.

Procedure

On the Cloudera Manager Server host, type the following commands to install the Cloudera Manager packages:

OS	Command
RHEL	<pre>sudo yum install cloudera-manager-daemons cloudera-manag er-agent cloudera-manager-server</pre>

Manually Install Cloudera Manager Agent Packages

The Cloudera Manager Agent is responsible for starting and stopping processes, unpacking configurations, triggering installations, and monitoring all hosts in a cluster. You can install the Cloudera Manager agent manually on all hosts, or Cloudera Manager can install the Agents in a later step. To use Cloudera Manager to install the agents, skip this section.

About this task

To install the Cloudera Manager Agent packages manually, do the following on every cluster host (including those that will run one or more of the Cloudera Management Service roles: Service Monitor, Activity Monitor, Event Server, Alert Publisher, or Reports Manager):

Procedure

1. Use one of the following commands to install the Cloudera Manager Agent packages:

OS	Command
RHEL, if you have a yum repo configured:	<pre>\$ sudo yum install cloudera-manager-agent cloudera-manager-daemons</pre>
RHEL, if you're manually transferring RPMs:	<pre>\$ sudo yum --nogpgcheck localinstall cloudera-manager-agent-package.*.x86_64.rpm cloudera-manager-daemons.*.x86_64.rpm</pre>

2. On every cluster host, 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
<code>server_host</code>	Name of the host where Cloudera Manager Server is running.
<code>server_port</code>	Port on the host where Cloudera Manager Server is running.

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

RHEL 7

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

If the agent starts without errors, no response displays.

When the Agent starts, it contacts the Cloudera Manager Server. If communication fails between a Cloudera Manager Agent and Cloudera Manager Server, see *Troubleshooting Installation Problems*. When the Agent hosts reboot, `cloudera-scm-agent` starts automatically.

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 [Overview of 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 [Configuring a Local Parcel Repository](#) on page 77.

Configuring a Local Parcel Repository

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

Related Information

[Overview of Parcels](#)

Using an Internally Hosted Remote Parcel Repository

The following sections describe how to use an internal Web server to host a parcel repository.

Related Information

[Overview of Parcels](#)

Setting Up a Web Server

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.

About this task

The examples on this page use Apache HTTP Server as the Web server. If you already have a Web server in your organization, you can skip to *Downloading and Publishing the Parcel Repository*.

Procedure

1. Install Apache HTTP Server:

RHEL / CentOS

```
sudo yum install httpd
```

2. Edit the Apache HTTP Server configuration file (/etc/httpd/conf/httpd.conf by default) to add or edit the following line in the <IfModule mime_module> section:

```
AddType application/x-gzip .gz .tgz .parcel
```

If the <IfModule mime_module> section does not exist, you can add it in its entirety as follows:



Note: This example configuration was modified from the default configuration provided after installing Apache HTTP Server on RHEL 7.

```
<IfModule mime_module>
#
# TypesConfig points to the file containing the list of mappings from
# filename extension to MIME-type.
#
TypesConfig /etc/mime.types
#
# AddType allows you to add to or override the MIME configuration
# file specified in TypesConfig for specific file types.
#
#AddType application/x-gzip .tgz
#
# AddEncoding allows you to have certain browsers uncompress
# information on the fly. Note: Not all browsers support this.
#
#AddEncoding x-compress .Z
#AddEncoding x-gzip .gz .tgz
#
# If the AddEncoding directives above are commented-out, then you
# probably should define those extensions to indicate media types:
#
AddType application/x-compress .Z
AddType application/x-gzip .gz .tgz .parcel

#
# AddHandler allows you to map certain file extensions to "handlers":
# actions unrelated to filetype. These can be either built into the se
rver
# or added with the Action directive (see below)
```

```
#
# To use CGI scripts outside of ScriptAliased directories:
# (You will also need to add "ExecCGI" to the "Options" directive.)
#
#AddHandler cgi-script .cgi

# For type maps (negotiated resources):
#AddHandler type-map var

#
# Filters allow you to process content before it is sent to the client
.
#
# To parse .shtml files for server-side includes (SSI):
# (You will also need to add "Includes" to the "Options" directive.)
#
AddType text/html .shtml
AddOutputFilter INCLUDES .shtml
</IfModule>
```



Warning: Skipping this step could result in an error message Hash verification failed when trying to download the parcel from a local repository, especially in Cloudera Manager 6 and higher.

3. Start Apache HTTP Server:

RHEL 7

```
sudo systemctl start httpd
```

Downloading and Publishing the Parcel Repository

Download the parcels that you want to install and publish the parcel directory.

Procedure

1. Download manifest.json and the parcel files for the product you want to install:

Runtime 7

Apache Impala, Apache Kudu, Apache Spark 2, and Cloudera Search are included in the Runtime parcel. To download the files for the latest Runtime 7 release, run the following commands on the Web server host:

```
sudo mkdir -p /var/www/html/cloudera-repos
sudo wget --recursive --no-parent --no-host-directories https://
[username]:[password]@archive.cloudera.com/p/cdh7/7.0.3.0/
parcels/ -P /var/www/html/cloudera-repos

sudo chmod -R ugo+rX /var/www/html/cloudera-repos/cdh7
```

Sqoop Connectors

To download the parcels for a Sqoop Connector release, run the following commands on the Web server host. This example uses the latest available Sqoop Connectors:

```
sudo mkdir -p /var/www/html/cloudera-repos
sudo wget --recursive --no-parent --no-host-directories http://ar
chive.cloudera.com/sqoop-connectors/parcels/latest/ -P /var/www/
html/cloudera-repos
```

```
sudo chmod -R ugo+rX /var/www/html/cloudera-repos/sqoop-connectors
```

If you want to create a repository for a different Sqoop Connector release, replace latest with the Sqoop Connector version that you want. You can see a list of versions in the parcels parent directory.

2. Visit the Repository URL `http://<Web_server>/cloudera-repos/` in your browser and verify the files you downloaded are present. If you do not see anything, your Web server may have been configured to not show indexes.

Related Information

[Overview of Parcels](#)

Configuring Cloudera Manager to Use an Internal Remote Parcel Repository

In Cloudera Manager's parcel settings, add a path to the internal parcel repository.

Procedure

1. Use one of the following methods to open the parcel settings page:
 - Navigation bar:
 - a. Click the parcel icon in the left navigation bar or click Hosts and click the Parcels tab.
 - b. Click the Configuration button.
 - Menu:
 - a. Select AdministrationSettings.
 - b. Select CategoryParcels.
2. Enter the path to the parcel. For example: `http://<web_server>/cloudera-parcels/cdh7/7.0.3.1/`

Using a Local Parcel Repository

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

Procedure

1. Open the Cloudera Manager Admin Console and click Parcels in the left-side navigation menu.
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 that 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. For instructions on downloading parcels, see [Downloading and Publishing the Parcel Repository](#) above.
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-6.1.0-1.cdh6.1.0.p0.770702-el7.parcel`:

```
shasum CDH-6.1.0-1.cdh6.1.0.p0.770702-el7.parcel | awk '{ print $1 }'  
> CDH-6.1.0-1.cdh6.1.0.p0.770702-el7.parcel.sha
```

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

```
sudo chown -R cloudera-scm:cloudera-scm /opt/cloudera/parcel-repo/*
```

8. In the Cloudera Manager Admin Console, click Parcels page in the left-side navigation menu.
9. Click Check for New Parcels and verify that the new parcel appears.

10. Download, distribute, and activate the parcel.

Production Installation: Installing Cloudera Manager, Cloudera Runtime, and Managed Services

This procedure is recommended for installing Cloudera Manager and Cloudera Runtime for production environments. For a non-production trial install see *Installing the CDP Private Cloud Base Trial*.

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

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. These tools depend on access to repositories to install software. Cloudera maintains Internet-accessible repositories for Runtime and Cloudera Manager installation files.

You can also create your own internal repository for hosts that do not have Internet access. For more information on creating an internal repository for Cloudera Manager, see *Configuring a Local Package Repository*.

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.

You can download the repository file at the following location:

```
https://[username]:[password]@archive.cloudera.com/p/cm7/7.0.3/redhat7/yum/cloudera-manager.repo
```

2. Import the repository signing GPG key:

- RHEL 7 compatible:

```
sudo rpm --import https://[username]:[password]@archive.cloudera.com/p/cm7/7.0.3/redhat7/yum/RPM-GPG-KEY-cloudera
```

3. Edit the repository file and add your username and password:

Open the /etc/yum.repos.d/cloudera-manager.repo file in a text editor. The file will look like this:

```
[cloudera-manager]
name=Cloudera Manager 7.0.3
baseurl=https://archive.cloudera.com/p/cm7/7.0.3/redhat7/yum/
gpgkey=https://archive.cloudera.com/p/cm7/7.0.3/redhat7/yum/RPM-GPG-KEY-cloudera
username=changeme
password=changeme
gpgcheck=1
enabled=1
autorefresh=0
type=rpm-md
```

and replace the two *changeme* placeholders with your username and password.

4. Continue to *Step 2: Install Java Development Kit*.

Step 2: Install Java Development Kit

CDP Private Cloud Base requires a JDK installed on all hosts., you can either install OpenJDK or a Oracle JDK directly from Oracle.

There are several options for installing a JDK on your CDP Private Cloud Base hosts:

- Install OpenJDK 8 on the Cloudera Manager server host and then allow Cloudera Manager to install OpenJDK 8 on its managed hosts.
- Manually install a [supported JDK](#) on all cluster hosts before installing Cloudera software.

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 .
- 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 operating system supported by CDP Private Cloud Base 7 uses 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 (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.

Installing OpenJDK Using Cloudera Manager

After you configure a repository, you can install the OpenJDK on the Cloudera Manager Server host using your package manager.

- RHEL Compatible

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

You can use Cloudera Manager to install the JDK on the remaining cluster hosts in an upcoming step. Continue to *Step 3. Installing Cloudera Manager Server*.

Manually Installing OpenJDK

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

About this task

OpenJDK 8 is the only supported OpenJDK version for CDP Data Center Edition.

Note that the path for the default truststore for OpenJDK 8 is `jre/lib/security/cacerts`.



Important: When you install CDP Private Cloud Base , Cloudera Manager includes an option to install Oracle JDK. De-select this option before continuing with the installation.

You must install a supported version of OpenJDK. If your deployment uses a version of OpenJDK lower than 1.8.0_181, see *TLS Protocol Error with OpenJDK*.



Note: If you intend to enable Auto-TLS, note the following:

You can specify a PEM file containing trusted CA certificates to be imported into the Auto-TLS truststore. If you want to use the certificates in the cacerts truststore that comes with OpenJDK, you must convert the truststore to PEM format first. However, OpenJDK ships with some intermediate certificates that cannot be imported into the Auto-TLS truststore. You must remove these certificates from the PEM file before importing the PEM file into the Auto-TLS truststore. This is not required when upgrading to OpenJDK from a cluster where Auto-TLS has already been enabled.

Procedure

Log in to each host and run the command for the version of the JDK you want to install:

RHEL

OpenJDK 8

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

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.

Procedure

1. Download the .tar.gz file for one of the 64-bit supported versions of the Oracle JDK from Java SE 8 Downloads.



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.

Results

After you have finished, continue to *Step 3: Install Cloudera Manager Server*.

Step 3: Install Cloudera Manager Server

In this step you install the Cloudera Manager packages on the Cloudera Manager Server host, and optionally enable auto-TLS.

Install Cloudera Manager Packages

Cloudera Manager is installed on the Cloudera Manager Server host using packages.

Procedure

1. On the Cloudera Manager Server host, type the following commands to install the Cloudera Manager packages:

OS	Command
RHEL	<pre>sudo yum install cloudera-manager-daemons cloudera-manager-agent cloudera-manager-server</pre>

2. If you are installing on Ubuntu, and are planning to add the Kudu service to the cluster and are planning to enable Apache Ranger, run the following command on all cluster hosts:

```
sudo apt-get install gettext-base
```



Note: If you know in advance which hosts will be running the Kudu service roles, you only need to run this command on those hosts.

(Recommended) Enable Auto-TLS

Auto-TLS greatly simplifies the process of enabling and managing TLS encryption on your cluster.



Note: Auto-TLS supports two options:

- Option 1: Use Cloudera Manager to generate an internal Certificate Authority and corresponding certificates
- Option 2: Use an existing Certificate Authority and corresponding certificates

The following procedure demonstrates Option 1, enabling auto-TLS to use an internal certificate authority (CA) created and managed by Cloudera Manager. To use a trusted public CA (Option 2), you must first obtain the certificates for your cluster hosts.

For new installations only, you can make the Cloudera Manager CA an intermediate CA to an existing internal root CA.

Auto-TLS automates the creation of an internal certificate authority (CA) and deployment of certificates across all cluster hosts. It can also automate the distribution of existing certificates, such as those signed by a public CA. Adding new cluster hosts or services to a cluster with auto-TLS enabled automatically creates and deploys the required certificates.

You can enable auto-TLS on existing clusters. If you do not want to enable auto-TLS right now, skip this section and continue to Step 4: Install and Configure Databases. Enabling auto-TLS on existing clusters is not supported if you are using the Cloudera Manager CA as an intermediate CA to an existing internal root CA, so if you want to use this option, you must enable auto-TLS now using the procedure documented in *Enabling Auto-TLS with an Existing Root CA*.

To enable auto-TLS with an embedded Cloudera Manager CA, run the following command:

```
sudo JAVA_HOME=/usr/java/jdk1.8.0_181-cloudera /opt/cloudera/cm-agent/bin/certmanager setup --configure-services
```



Note: The certmanager utility is included with Cloudera Manager Agent, but not Cloudera Manager Server. If you see an error about the certmanager command not being found, make sure you have installed the cloudera-manager-agent package as documented above.

Replace *jdk1.8.0_181-cloudera* with your JDK version. If you want to store the files in a directory other than the default (*/var/lib/cloudera-scm-server/certmanager*), add the *--location* option as follows:

```
sudo JAVA_HOME=/usr/java/jdk1.8.0_181-cloudera /opt/cloudera/cm-agent/bin/certmanager --location /opt/cloudera/CMCA setup --configure-services
```

Check the */var/log/cloudera-scm-agent/certmanager.log* log file to confirm that the */var/lib/cloudera-scm-server/certmanager/** directories were created.

When you start Cloudera Manager Server, it will have TLS enabled, and all hosts that you add to the cluster, as well as any supported services, will automatically have TLS configured and enabled.

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.

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 database for the Cloudera Manager Server and other services that use databases.

Required Databases

The following components all require databases: Cloudera Manager Server, Oozie Server, Sqoop Server, Reports Manager, Hive Metastore Server, Hue Server, DAS server, and Ranger.

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.

- 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.
- 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.
- DAS PostgreSQL server - Contains Hive and Tez event logs and DAG information. Can grow very large.
- Ranger Audit - Apache Ranger uses Apache Solr to store audit logs and provides UI searching through the audit logs. The size of the audit database depends on the events in the environment. You should monitor and manage the auditing destinations. 1TB of space is recommended.



Note: Solr must be installed and configured before installing Ranger Admin or any of the Ranger component plugins.

- Ranger Admin - Contains administrative information such as Ranger users, groups, and access policies. Medium-sized.

The Host Monitor and Service Monitor services use local disk-based datastores.

The JDBC connector for your database must be installed on the host where you assign the Reports Manager role.

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

Related Information

[Step 2. Configure the Ranger and DAS Databases](#)

Install and Configure PostgreSQL for CDP

To use a PostgreSQL database, follow these procedures. For information on compatible versions of the PostgreSQL database, see *Database Requirements*.



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 trial installations.

Installing PostgreSQL Server

Install the PostgreSQL packages on the PostgreSQL server.



Note:

- If you already have a PostgreSQL database set up, you can skip to the section *Configuring and Starting the PostgreSQL Server* 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 Runtime component databases (such as Hive and Hue). For more information, see *Schemas* in the PostgreSQL documentation.

Install the PostgreSQL packages as follows:

RHEL:

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

Installing the psycopg2 Python Package

Hue in Runtime 7 requires version 2.7.5 or higher of the psycopg2 Python package for connecting to a PostgreSQL database. The psycopg2 package is automatically installed as a dependency of Cloudera Manager Agent, but the version installed is often lower than 2.7.5.

If you are installing Runtime 7 and using PostgreSQL for the Hue database, you must install psycopg2 2.7.5 or higher on all Hue hosts as follows. These examples install version 2.7.5:

RHEL 7 Compatible

1. Install the python-pip package:

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

2. Install psycopg2 2.7.5 using pip:

```
sudo pip install psycopg2==2.7.5 --ignore-installed
```

Configuring and Starting the PostgreSQL Server

By default, PostgreSQL only accepts connections on the loopback interface. Configure PostgreSQL to accept the connections based on hostname, IP address (including CIDR address), or MAC address. A fully qualified domain name (FQDN) is not a requirement. If you do not make these changes, the services cannot connect to and use the database on which they depend.

Before you begin

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

Procedure

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"
```

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 <range-start-ip-address>/28 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
 - `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.
 - `checkpoint_completion_target` - 0.9.
4. Configure the PostgreSQL server to start at boot.

OS	Command
RHEL 7 compatible	<pre>sudo systemctl enable postgresql</pre>

5. Restart the PostgreSQL database:

- RHEL 7 Compatible:

```
sudo systemctl restart postgresql
```

Creating Databases for CDP

You must create databases and service accounts for components that require databases.

About this task

The following components require databases:

- Cloudera Manager Server
- Cloudera Management Service roles:
- Reports Manager
- Hue
- Each Hive metastore
- Oozie
- Data Analytics Studio
- Ranger

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.



Note: The instructions for Cloudera Manager Server, Cloudera Management Service roles, Activity Monitor, Reports Manager, Hue, Hive metastores, Oozie, and DAS are documented in this topic.

Additional configuration for Ranger is documented in the following two topics. Refer to those topics for detailed instructions on the Ranger database.



Note: To use DAS, install the PostgreSQL database version 9.6.

To create databases for Cloudera Manager Server, Cloudera Management Service roles, Activity Monitor, Reports Manager, Hue, Hive metastores, Oozie, and DAS, complete the following steps:

Procedure

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 19: 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

Service	Database	User
Oozie	oozie	oozie
Data Analytics Studio	das	das
Ranger	ranger	rangeradmin

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

What to do next

If you plan to use Apache Ranger, see the following topic for instructions on creating and configuring the Ranger database.

After you install and configure PostgreSQL databases for Cloudera software, continue to *Step 5: Set up the Cloudera Manager Database* to configure a database for Cloudera Manager.

Related Information

[Step 2. Configure the Ranger and DAS Databases](#)

Configuring a PostgreSQL Database for Ranger

Complete the following steps to configure a PostgreSQL database instance for Ranger.

Configuring a PostgreSQL Database for Ranger on RHEL7/Centos7

Procedure

1. Run the following command to install PostgreSQL server:

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

2. Initialize the Postgres database and start PostgreSQL:

```
sudo postgresql-setup initdb
sudo systemctl start postgresql
```

3. Optional: Configure PostgreSQL to start on boot:

```
sudo systemctl enable postgresql
```

4. Update the postgresql.conf file, which is usually found in /var/lib/pgsql/data or /var/lib/postgresql/data:

- Uncomment and change #listen_addresses = 'localhost' to listen_addresses = '*'
- Uncomment the #port = line and specify the port number (the default is 5432)
- Optional: Uncomment and change #standard_conforming_strings= to standard_conforming_strings = off

5. Update the pg_hba.conf file, which is usually found in /var/lib/pgsql/data or /etc/postgresql/<version>/main:

- Add the following line to allow connection to the Ranger database from any host:

```
host    ranger          rangeradmin    0.0.0.0/0          md5
```

6. Restart PostgreSQL:

```
sudo systemctl restart postgresql
```

7. The PostgreSQL database administrator should be used to create the Ranger databases. The following series of commands could be used to create the rangeradmin user and grant it adequate privileges. Be sure to replace 'password' with a strong password.

```
echo "CREATE DATABASE ranger;" | sudo -u postgres psql -U postgres
echo "CREATE USER rangeradmin WITH PASSWORD 'password';" | sudo -u postgres psql -U postgres
```

```
echo "GRANT ALL PRIVILEGES ON DATABASE ranger TO rangeradmin;" | sudo -u postgres psql -U postgres
```

8. Install the PostgreSQL JDBC driver. If you would like to use the PostgreSQL JDBC driver version shipped with the OS repositories, run the following command:

```
yum install postgresql-jdbc*
```

You can also download the JDBC driver from the official PostgreSQL JDBC Driver website – <https://jdbc.postgresql.org/>.

9. Rename the Postgres JDBC driver .jar file to postgresql-connector-java.jar and copy it to the /usr/share/java directory. The following copy command can be used if the Postgres JDBC driver .jar file is installed from the OS repositories:

```
cp /usr/share/java/postgresql-jdbc.jar /usr/share/java/postgresql-connector-java.jar
```

10. Confirm that the .jar file is in the Java share directory:

```
ls /usr/share/java/postgresql-connector-java.jar
```

11. Change the access mode of the .jar file to 644:

```
chmod 644 /usr/share/java/postgresql-connector-java.jar
```

What to do next

Ensure that the Ranger Solr and Ranger HDFS plugins are enabled. See [Additional Steps for Apache Ranger](#) on page 98 for details.

Step 5: Set up and configure the Cloudera Manager database

Cloudera Manager Server includes the scm_prepare_database.sh script that can create and configure a database.

The scm_prepare_database.sh script can perform the following activities::

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

The scm_prepare_database.sh script checks the connection between the Cloudera Manager Server and the database. Upon successful connection, the script writes the /etc/cloudera-scm-server/db.properties file. When you start Cloudera Manager for the first time, the scm_prepare_database.sh script creates and populates the necessary tables.

Although the script can create a database, the following procedures assume that you have already created the database as described in *Install and Configure Databases*. For more information about tuning the Cloudera Manager database for best performance, see the corresponding Knowledge article: [hibernate.c3p0 Configs for Cloudera Manager](#).

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

Syntax for scm_prepare_database.sh

Review the syntax of the scm_prepare_database.sh script before you run it to configure the Cloudera Manager database.

The syntax for the scm_prepare_database.sh script is as follows:

```
sudo /opt/cloudera/cm/schema/scm_prepare_database.sh [option s] <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*, do not specify these options.

The following tables describe the parameters and options for the `scm_prepare_database.sh` script:

Table 20: Parameters

Parameter (Required in bold)	Description
<code><databaseType></code>	One of the supported database types: <ul style="list-style-type: none"> PostgreSQL: <code>postgresql</code>
<code><databaseName></code>	The name of the Cloudera Manager Server database to use. For PostgreSQL databases, the script can create the specified database if you specify the <code>-u</code> and <code>-p</code> 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 <code>scm</code> , but you can also use any other database name such as <code>cm_db</code> or <code>cmdb1</code> .
<code><databaseUser></code>	The username for the Cloudera Manager Server database to create or use. The default username provided in the Cloudera Manager configuration settings is <code>scm_user</code> , but you can also use any other database user such as <code>cm_user</code> or <code>cm_db_user</code> .
<code><password></code>	The password for the <code><databaseUser></code> 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: <div style="background-color: #f0f0f0; padding: 10px; margin-top: 10px;"> Enter SCM password: </div>

Table 21: Options

Option	Description
<code>-? --help</code>	Display help.
<code>--config-path</code>	The path to the Cloudera Manager Server configuration files. The default is <code>/etc/cloudera-scm-server</code> .
<code>-f --force</code>	If specified, the script does not stop if an error occurs.
<code>-h --host</code>	The IP address or hostname of the host where the database is installed. The default is to use <code>localhost</code> .
<code>-p --password</code>	The admin password for the database application. Use with the <code>-u</code> option. The default is no password. Do not put a space between <code>-p</code> and the password (for example, <code>-phunter2</code>). If you do not want the password visible on the screen or stored in the command history, use the <code>-p</code> option without specifying a password, and you are prompted to enter it as follows: <div style="background-color: #f0f0f0; padding: 10px; margin-top: 10px;"> Enter database password: </div> If you have already created the database, do not use this option.
<code>-P --port</code>	The port number to use to connect to the database. The default port is: <ul style="list-style-type: none"> PostgreSQL: <code>5432</code> This option is used for a remote connection only.
<code>--scm-host</code>	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 <code>-h</code> option.
<code>--scm-password-script</code>	A script to execute whose stdout provides the password for user SCM (for the database).
<code>-u --user</code>	The admin username for the database application. Use with the <code>-p</code> option. Do not put a space between <code>-u</code> and the username (for example, <code>-uroot</code>). 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.

Step 6: Install Runtime and Other Software

After you set up the Cloudera Manager database, start Cloudera Manager Server and log in to the Cloudera Manager Admin Console. Then proceed through the installation wizard.

Procedure

1. Start Cloudera Manager Server:

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

2. If you want to configure the Cloudera Manager server to start automatically when the host reboots, run the following command:

```
sudo systemctl enable cloudera-scm-server
```

3. 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*.

4. 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.



Note: If you enabled auto-TLS, you are redirected to `https://<server_host>:7183`, and a security warning is displayed. You might need to indicate that you trust the certificate, or click to proceed to the Cloudera Manager Server host.

5. 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.

Results

After logging in, the installation wizard launches. The following sections guide you through each step of the installation wizard.

Installation Wizard

Proceed through the installation wizard to accept licenses, install and configure Cloudera Runtime, and more.

Upload License File

On the Upload License File page, you can select either the trial version of CDP Data Center or upload a license file:

1. Choose one of the following options:
 - Upload Cloudera Data Platform License
 - Try Cloudera Data Platform for 60 days. The CDP Data Center trial does not require a license file, but the trial expires after 60 days.
2. If you choose the CDP Data Center Edition Trial, you can upload a license file at a later time. Read the license agreement and click the checkbox labeled Yes, I accept the Cloudera Standard License Terms and Conditions if you accept the terms and conditions of the license agreement. Then click Continue.
3. If you have a license file for CDP Data Center, upload the license file:
 - a. Select Upload Cloudera Data Platform License.
 - b. Click Upload License File.
 - c. Browse to the location of the license file, select the file, and click Open.
 - d. Click Upload.
 - e. Click Continue.
4. Information is displayed indicating what the Runtime installation includes. At this point, you can click the Support drop-down menu to access online Help or the Support Portal.
5. Click Continue to proceed with the installation.

Welcome (Add Cluster - Installation)

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

Click Continue to proceed with the installation.

Cluster Basics

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

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

For new installations, Regular Cluster is the only option. You cannot add a compute cluster if you do not have an existing base cluster.

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

Enter a cluster name and then click Continue.

Specify Hosts

Choose which hosts will run Runtime and other managed services.



Note: If you have enabled Auto-TLS, you must include the Cloudera Manager server host when you specify hosts.

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

Expansion Range	Matching Hosts
10.1.1.[1-4]	10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4
host[1-3].example.com	host1.example.com, host2.example.com, host3.example.com

Expansion Range	Matching Hosts
host[07-10].example.com	host07.example.com, host08.example.com, host09.example.com, host10.example.com



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

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

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



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

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

The Select Repository screen displays.

Select Repository

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

In the Cloudera Manager Agent section:

1. Select either Public Cloudera Repository or Custom Repository for the Cloudera Manager Agent software.
2. If you select Custom Repository, do not include the operating system-specific paths in the URL. For instructions on setting up a custom repository, see *Configuring a Local Package Repository*.

In the CDH and other software section:

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

A parcel is a binary distribution format containing the program files, along with additional metadata used by Cloudera Manager. Parcels are required for rolling upgrades. For more information, see *Parcels*.
 - Packages are not supported for Runtime in Data Center 7.0.
2. Select the version of Runtime to install. If 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 Runtime 7 are documented in *Cloudera Runtime 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.



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.

3. If you selected Use Parcels, specify any Additional Parcels you want to install.
4. Click Continue.

Select JDK



Note: CDP Data Center is no longer bundled with Oracle JDK software. Cloudera provides a supported version of OpenJDK.

If you installed your own JDK version, such as Oracle JDK 8, in *Step 2: Install Java Development Kit*, select **Manually manage JDK**.

To allow Cloudera Manager to automatically install the OpenJDK on cluster hosts, select **Install a Cloudera-provided version of OpenJDK**.

To install the default OpenJDK that is provided by your operating system, select **Install a system-provided version of OpenJDK**.

After checking the applicable boxes, click **Continue**.

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 Manager Agent Packages* for the procedure and then continue with the next steps on this page.

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

If you are using parcels, the **Install Parcels** page displays. If you chose to install using packages, the **Inspect Cluster** 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 Cluster** page displays.

Inspect Cluster

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

1. Run the Inspect Network Performance tool. You can click Advanced Options to customize some ping parameters.
2. After the network inspector completes, click Show Inspector Results to view the results in a new tab.
3. Address any reported issues, and click Run Again (if applicable).
4. Click Inspect Hosts to run the Host Inspector utility.
5. After the host inspector completes, click Show Inspector Results to view the results in a new tab.
6. Address any reported issues, and click Run Again (if applicable).

If the reported issues cannot be resolved in a timely manner, and you want to abandon the cluster creation wizard to address them, select the radio button labeled **Quit the wizard** and Cloudera Manager will delete the temporarily created cluster and then click **Continue**.

Otherwise, after addressing any identified problems, select the radio button labeled **I understand the risks, let me continue with cluster creation**, and then click **Continue**.

This completes the Cluster Installation wizard and launches the Add Cluster - Configuration wizard.

Continue to *Step 7: Set Up a Cluster Using the Wizard*.

Step 7: Set Up a Cluster Using the Wizard

After you complete the Add Cluster - Installation wizard, the Add Cluster - Configuration 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.

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

Choose one of the following:

Regular (Base) Clusters

Data Engineering

Process develop, and serve predictive models.

Services included: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue, and Data Analytics Studio

Data Mart

Browse, query, and explore your data in an interactive way.

Services included: HDFS, Ranger, Atlas, Hive, and Hue

Operational Database

Real-time insights for modern data-driven business.

Services included: HDFS, Ranger, Atlas, and HBase

Custom Services

Choose your own services. Services required by chosen services will automatically be included.

Compute Clusters

Data Engineering

Process develop, and serve predictive models.

Services included: Spark, Oozie, Hive on Tez, Data Analytics Studio, HDFS, YARN, and YARN Queue Manager

Spark

Spark for Compute

Services included: Core Configuration, Spark, Oozie, YARN, and YARN Queue Manager

Streams Messaging (Simple)

Simple Kafka cluster for streams messaging

Services included: Kafka, Schema Registry, and Zookeeper

Streams Messaging (Full)

Advanced Kafka cluster with monitoring and replication services for streams messaging

Services included: Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control, and Zookeeper

Custom Services

Choose your own services. Services required by chosen services will automatically be included.

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.

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 hosts, names, usernames, and passwords you created in *Step 4: Install and Configure Databases*.

For services that support it, you can add finer-grained customizations using a JDBC URL override.



Important: The Hive service is currently the only service that supports the JDBC URL override.

Select the database type and enter the database name, username, and password for each service.

For services that support it, to specify a JDBC URL override, select Yes in the Use JDBC URL Override dropdown menu. You must also specify the database type, username, and password.

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.

Enter Required Parameters

The **Enter Required Parameters** page lists required parameters for DAS, the Cloudera Manager API client, and Ranger.

The DAS database hostname, database name, database username, and database password were configured when you created the required DAS database. The default database name is “das” and the default database user is “das”.

If you do not have an existing user for the Cloudera Manager API client, use the default username and password “admin” for both the The Existing Cloudera Manager API Client Username and The Existing Cloudera Manager API Client Password.

The Atlas Admin user, Ranger Admin user, Usersync user, Tagsync user, and KMS Keyadmin user are created during cluster deployment. In this page you must give a password for each of these users.



Note: Passwords for the Atlas Admin, Ranger Admin, Usersync, Tagsync, and KMS Keyadmin users must be a minimum of 8 characters long, with at least one alphabetic and one numeric character. The following characters are not valid: " ' \ ` ' .

The Ranger database host, name, user, and user password were configured when you created the required Ranger database. If you ran the `gen_embedded_ranger_db.sh` script to create the Ranger database, the output of the script contained the host and database user password. Enter those here. The default database name is "ranger" and the default database user is "rangeradmin."

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 Command Details page displays.

Command Details

The Command Details 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 Only Running Steps.

After the First Run command completes, click Continue to go to the Summary page.

If cluster deployment fails, be sure to click Resume in the wizard after you fix any issues. If you do not click Resume, the Ranger service will not enable all of the necessary plugins.

Summary

The Summary page reports the success or failure 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.

Additional Steps for Apache Ranger

The Ranger plugins for HDFS and Solr may not be enabled by default. Ranger plugins enable Cloudera Manager stack components – such as HDFS and Solr – to connect to Ranger and access its authorization and audit services. Verify that the HDFS and Solr plugins are enabled after you install and start the Ranger service.

Procedure

1. To enable the HDFS plugin:
 - a) Login to Cloudera Manager.
 - b) Go to the HDFS Service status page.
 - c) Click the Configuration tab.
 - d) Search for the Enable Ranger Authorization configuration property.
 - e) If the Enable Ranger Authorization property is not selected, select it and save the changes.
 - f) Go to the Ranger Service status page and click ActionsSetup Ranger Plugin Service.
 - g) Restart the HDFS service.

2. To enable the Ranger Solr plugin:
 - a) Login to Cloudera Manager.
 - b) Go to the Solr Service status page.
 - c) Click the Configuration tab.
 - d) Search for the Enable Ranger Authorization configuration property.
 - e) If the Enable Ranger Authorization property is not selected, select it and save the changes.



Note: Don't select the Ranger Service dependency parameter. This is used for enabling a Solr service instance that is not used by the Ranger service.

- f) Restart the Solr service.

Installing Cloudera Navigator Key Trustee Server

You can install Navigator Key Trustee Server using Cloudera Manager with parcels. Command line package installs are not supported.



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.



Note: If you are using or planning to use Key Trustee Server in conjunction with a Runtime 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.

See *Encrypting Data at Rest* for more information about encryption and Key Trustee Server requirements.

Installing Key Trustee Server Using Cloudera Manager

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.

Procedure

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*.



Important: The Add Cluster wizard prompts you to install Runtime and other cluster services. To exit the wizard without installing Runtime, select a version of Runtime 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 Runtime or other services that are not required for Key Trustee Server.

2. In Cloudera Manager, go to HostsParcels.
3. Click Configuration and add the path to the Key Trustee Server parcel to the Remote Parcel Repository URLs section.

Key Trustee Server Version	Parcel Repository URL
7.0.3.0	https://archive.cloudera.com/p/keytrusteeserver7/7.0.3.0/

- 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.

What to do next

After installing Key Trustee Server using Cloudera Manager, continue to *Securing Key Trustee Server Host*.

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 7.

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.

```
# 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. See the Key Trustee Server documentation and High Availability documentation for details.

Related Information

[Initializing Standalone Key Trustee Server](#)

[Cloudera Navigator Key Trustee Server Overview](#)

[Setting Up Key Trustee Server High Availability](#)

Installing Key Trustee KMS

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. When you install Key Trustee KMS, you add the parcel repository URL in Cloudera Manager and then download, distribute, and activate the parcel.



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* 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 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, which uses 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.

Installing Key Trustee KMS Using Parcels

To install Key Trustee KMS, add the parcel repository URL in Cloudera Manager and then download, distribute, and activate the parcel.

Procedure

1. In Cloudera Manager, go to HostsParcels.
2. Click Configuration and add the path to the Key Trustee KMS parcel to the Remote Parcel Repository URLs section.

Key Trustee KMSVersion	Parcel Repository URL
7.0.3.0	https://archive.cloudera.com/p/keytrustee/7.0.3.0/

3. Download, distribute, and activate the Key Trustee KMS parcel.



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.

What to do next

For instructions on installing Key Trustee Server and configuring Key Trustee KMS to use Key Trustee Server, see the topics *Installing Cloudera Navigator Key Trustee Server* and *Enabling HDFS Encryption Using the Wizard*.

After You Install

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.

Testing the Installation

Begin testing the installation from the **Home** page, where you can start by checking the health of the services.

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.

Testing with Hue

You can test the cluster by running Hue.

About this task

Hue is a graphical user interface that allows you to interact with your clusters by running applications that let you browse HDFS and cloud object storage such as S3 and ABFS, manage a Hive metastore, and run Hive, Impala, and Search queries, and Oozie workflows.

Procedure

1. From Cloudera Manager, go to Clusters Hue service .
2. Click Web UI link and select the Hue web URL, which opens Hue in a new window.
By default, Authentication Backend is set to AllowFirstUserDjangoBackend. This makes the first user who logs into Hue the Superuser and allows you to set the username and password, and create other users.
You can change the Authentication Backend as per your requirements from Hue configurations in Cloudera Manager.
3. You can run a query or browse the database that you have set up for Hue.
For more information, see the Hue documentation.

Secure Your Cluster

After completing your Cloudera Enterprise installation and making sure that everything is working properly, secure your cluster by enabling authentication, authorization, auditing, and encryption.

For comprehensive instructions on securing your cluster, see the Security documentation.

Troubleshooting Installation Problems

This topic describes common installation issues and suggested solutions.

TLS Protocol Error with OpenJDK

If you are using an older version of OpenJDK 1.8 and have enabled SSL/TLS for the Cloudera Manager Admin Console, you may encounter a TLS protocol error when connecting to the Admin Console, stating that there are no ciphers in common. This is because older versions of OpenJDK may not implement certain TLS ciphers, causing an inability to log into the Cloudera Manager Admin Console when TLS is enabled.

Workaround:

You can workaround this issue by doing one of the following:

- Upgrade OpenJDK to a supported version of OpenJDK that is higher than version 1.8.0_181.
- If it is not possible to upgrade OpenJDK, enable less secure TLS ciphers in Cloudera Manager. You can do this by opening the /etc/default/cloudera-scm-server in a text editor and adding the following line:

```
export CMF_OVERRIDE_TLS_CIPHERS=<cipher_list>
```

Where <cipher_list> is a list of TLS cipher suites separated by colons. For example:

```
export CMF_OVERRIDE_TLS_CIPHERS="TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256:
TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256:TLS_ECDHE_ECDSA_WITH_AES_256_GCM_
SHA384:TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384:TLS_DHE_RSA_WITH_AES_128_GC
M_SHA256:TLS_DHE_RSA_WITH_AES_256_GCM_SHA384:TLS_ECDHE_ECDSA_WITH_AES_12
8_CBC_SHA256:TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256:TLS_ECDHE_ECDSA_WITH_
AES_128_CBC_SHA:TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384:TLS_ECDHE_RSA_WITH_
_AES_128_CBC_SHA:TLS_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384:TLS_ECDHE_ECDSA_
_WITH_AES_256_CBC_SHA:TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA:TLS_DHE_RSA_WIT
H_AES_128_CBC_SHA256:TLS_DHE_RSA_WITH_AES_128_CBC_SHA:TLS_DHE_RSA_WITH_A
ES_256_CBC_SHA256:TLS_DHE_RSA_WITH_AES_256_CBC_SHA:TLS_ECDHE_ECDSA_WITH_
3DES_EDE_CBC_SHA:TLS_ECDHE_RSA_WITH_3DES_EDE_CBC_SHA:TLS_EDH_RSA_WITH_3D
ES_EDE_CBC_SHA:TLS_RSA_WITH_AES_128_GCM_SHA256:TLS_RSA_WITH_AES_256_GCM_
SHA384:TLS_RSA_WITH_AES_128_CBC_SHA256:TLS_RSA_WITH_AES_256_CBC_SHA256:T
LS_RSA_WITH_AES_128_CBC_SHA:TLS_RSA_WITH_AES_256_CBC_SHA:TLS_RSA_WITH_3D
ES_EDE_CBC_SHA"
```


Cloudera Bug: OPSAPS-49578

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 reason:

You might have SELinux enabled.

Possible solution:

Disable SELinux by running `sudo setenforce 0` on the Cloudera Manager Server host. To disable it permanently, edit /etc/selinux/config.

Installation interrupted and installer does not restart

Possible reason:

You need to do some manual cleanup.

Possible solution:

See *Uninstalling Cloudera Manager and Managed Software*.

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 reason:

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 solution:

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;`

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 reason:

You might have SELinux or iptables enabled.

Possible solution:

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.

Cluster hosts do not appear

Some cluster hosts do not appear when you click Find Hosts in install or update wizard.

Possible reason:

You might have network connectivity problems.

Possible solution:

- 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.

"Access denied" in install or update wizard

"Access denied" in install or update wizard during database configuration for Reports Manager.

Possible reason:

Hostname mapping or permissions are not set up correctly.

Possible solution:

- For hostname configuration, see *Configure 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.

Reports Manager or Service Monitor databases fail to start.

Possible reason:

MySQL binlog format problem.

Possible solution:

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 85.

Cloudera services fail to start

Possible reason:

Java might not be installed or might be installed at a custom location.

Possible solution:

See *Configuring a Custom Java Home Location* for more information on resolving this issue.

Create Hive Metastore Database Tables command fails

The Create Hive Metastore Database Tables command fails due to a problem with an escape string.

Possible reason:

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 solution:

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 `AuditActivity` 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 reason:

You have not granted execute permission to `sys.dbms_crypto`.

Possible solution:

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

Complete the following tasks to uninstall the Cloudera Manager Server, Agents, managed software, and databases.

Record User Data Paths

Record the location of the user data paths by checking the configuration in each service.

The user data paths listed in the topic *Remove User Data*, `/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 Runtime and managed services or at some later time, note the location of the paths by checking the configuration in each service.

Stop all Services

Stop all services for each cluster managed by Cloudera Manager.

Procedure

1. On the HomeStatus tab, click three dots to the right of the cluster name and select Stop.
2. Click Stop in the confirmation screen. The Command Details window shows the progress of stopping services. When All services successfully stopped appears, the task is complete and you can close the Command Details window.
3. On the HomeStatus tab, click the three dots to the right of the Cloudera Management Service entry and select Stop. The Command Details window shows the progress of stopping services.

Results

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*; you will remove packages in *Uninstall Cloudera Manager Agent and Managed Software*. If you installed using parcels remove them as follows:

Procedure

1.



Click the parcel indicator in the left-hand navigation bar.

2. In the Location selector on the left, select All Clusters.
3. For each activated parcel, select ActionsDeactivate. When this action has completed, the parcel button changes to Activate.
4. For each activated parcel, select ActionsRemove from Hosts. When this action has completed, the parcel button changes to Distribute.
5. For each activated parcel, select ActionsDelete. This removes the parcel from the local parcel repository.

What to do next

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.

Procedure

1. If you used the cloudera-manager-installer.bin file (the trial installer): Run the following command on the Cloudera Manager Server host:
2. 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:
 - a) Stop the Cloudera Manager Server and its database:

```
sudo /opt/cloudera/installer/uninstall-cloudera-manager.sh
```

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

- b) 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-2` steps.

RHEL systems:

```
sudo yum remove cloudera-manager-server
sudo yum remove cloudera-manager-server-db-2
```

```
sudo zypper -n rm --force-resolution cloudera-manager-server
sudo zypper -n rm --force-resolution cloudera-manager-server-db-2
```

```
sudo apt-get remove cloudera-manager-server
sudo apt-get remove cloudera-manager-server-db-2
```

Uninstall Cloudera Manager Agent and Managed Software

To uninstall Cloudera Manager Agent and managed software, stop the Cloudera Manager Agent on all hosts, remove the parcel installation, and run the clean command.

About this task

Do the following on all Agent hosts:

Procedure

1. Stop the Cloudera Manager Agent.

RHEL 7, SLES 12, Debian 8, Ubuntu 16.04 and higher

```
sudo systemctl stop supervisord
```

2. To uninstall managed software, run the following commands:

RHEL: `$ sudo yum remove 'cloudera-manager-*`

3. Run the clean command:

RHEL

```
sudo yum clean all
```

```
sudo apt-get clean
```

Remove Cloudera Manager, User Data, and Databases

Permanently remove Cloudera Manager data, the Cloudera Manager lock file, and user data. Then stop and remove the databases.

Procedure

1. 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.

2. 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*
```

3. On all Agent hosts, run this command to remove the Cloudera Manager lock file:

```
sudo rm /tmp/.scm_prepare_node.lock
```

4. 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.

- a) 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
```

- b) 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
```

5. Stop and remove the 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 Runtime Component From a Single Host

The following procedure removes Runtime software components from a single host that is managed by Cloudera Manager.

Procedure

1. In the Cloudera Manager Administration Console, select HostsAll Hosts.
A list of hosts in the cluster displays.
2. Select the host where you want to uninstall Runtime software.
3. Click the Actions for Selected button and select Remove From Cluster.
Cloudera Manager removes the roles and host from the cluster.
4. Optionally, manually delete the krb5.conf file used by Cloudera Manager.