

# Getting Started with Cloudera Base on premises Upgrade and Migration

Date published: 2019-11-22

Date modified: 2024-12-10

**CLOUDERA**

# Legal Notice

© Cloudera Inc. 2026. All rights reserved.

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

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

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

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

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

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

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

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

# Contents

<b>Cloudera Upgrade and Migrations Paths.....</b>	<b>4</b>
<b>Regressions During an Upgrade.....</b>	<b>6</b>
<b>Downloadable Cloudera upgrade checklists.....</b>	<b>8</b>
<b>Supported in-place upgrade paths.....</b>	<b>9</b>
Cloudera Manager Support Matrix.....	17
<b>Cloudera Base on premises requirements and supported versions.....</b>	<b>21</b>
Hardware Requirements.....	21
Cloudera Manager.....	21
Cloudera Runtime.....	24
Operating System Requirements.....	35
Database Requirements.....	39
RDBMS High Availability Support.....	40
Java Requirements.....	42
Supported JDKs.....	42
Support Notes.....	44
Networking and Security Requirements.....	45
Data at Rest Encryption Requirements.....	50
Third-party filesystems.....	50
Third-party filesystem support: Dell EMC PowerScale.....	50
Third-party filesystem support: IBM Spectrum Scale.....	52
<b>Data Migration Versus Upgrade.....</b>	<b>53</b>

# Cloudera Upgrade and Migrations Paths

Take a look at the overview, features, and advantages of Cloudera and know the upgrade and migration paths from CDH or HDP platform to Cloudera.

## Introduction to Cloudera

The merger of Cloudera and Hortonworks led to the new Cloudera, which is the combined best of breed Big Data components from both Cloudera and Hortonworks.

Review the following information before you upgrade or migrate to Cloudera:

- Cloudera Overview
- Cloudera Base on premises new features

## Troubleshooting

A selection of Cloudera Knowledge Base articles are available that describe common issues encountered by Cloudera customers during upgrades and migrations. See [Cloudera Upgrade/Migrate Troubleshooting Articles](#). (Cloudera login required.)

## Cloudera upgrade and migration paths are:

### Important:



Before performing an upgrade of Cloudera Manager or the Cloudera Runtime, creating a backup of all the metadata databases is important. This includes the Cloudera Manager database, and the various Cloudera Runtime component databases such as Hive Metadata Server, Ranger Admin, Ranger KMS, Schema Registry, and so on. The backups are necessary if there is a reason to rollback to the prior version.

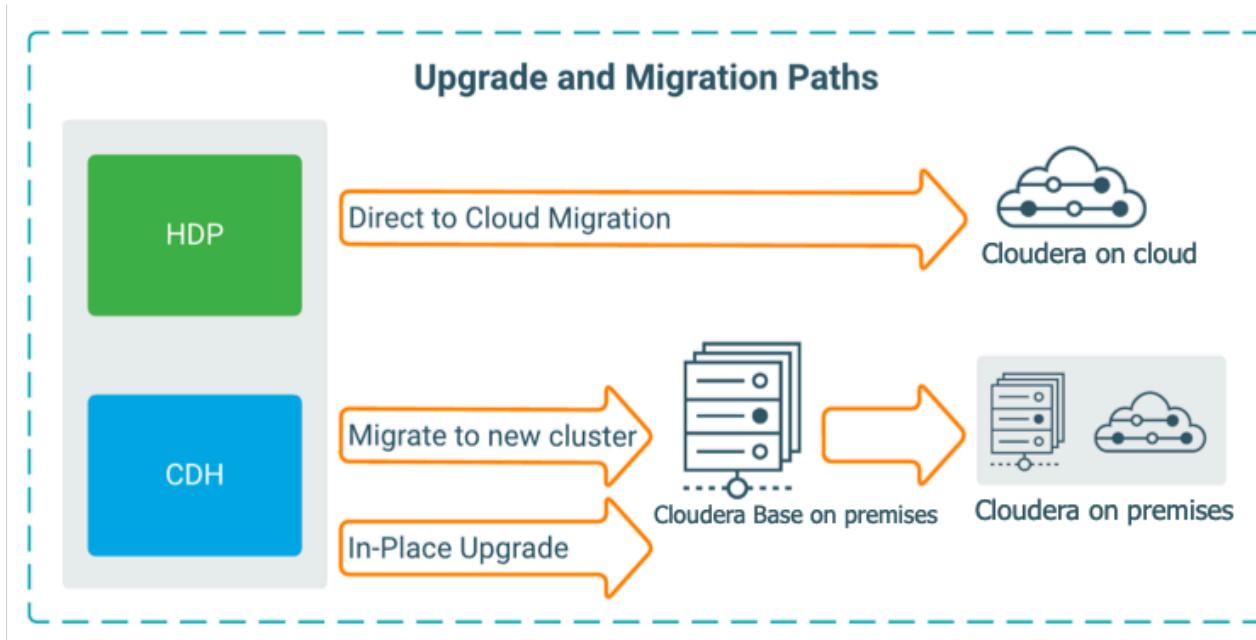
- [Cloudera Upgrade and Migration paths for Data-at-Rest](#)
- [Cloudera Upgrade and Migration paths to Cloudera for Cloudera DataFlow Functions components](#)
- [Cloudera Upgrade and Migration paths for Cloudera Data Science Workbench \(Cloudera AI\)](#)



**Note:** If you are upgrading to Cloudera Manager 7.5.1 or higher in order to install Cloudera Data Services on premises version 1.5.4, you must use Cloudera Runtime version 7.1.6 or higher. For more information, see [Cloudera Data Services on premises](#).

## Cloudera Upgrade and Migration Paths for Data-at-Rest

If you are an HDP or a CDH user, you can follow one of the several upgrade or migration paths to Cloudera.



Migration and Upgrade Paths	Description
In-place upgrade	<p>Recommended for large clusters. Other paths are not viable. Involves downtime. See the following to know more about the In-place upgrade paths:</p> <ul style="list-style-type: none"> <li><a href="#">In-place upgrade from CDH to CDP Private Cloud Base</a></li> <li><a href="#">In-place upgrade from HDP to CDP Private Cloud Base</a></li> </ul>
Migrate to new cluster	<p>Recommended if you are ready for a hardware refresh or have small clusters. You can fall back to your original cluster in the event of upgrade issues. Requires additional hardware.</p>
Direct to cloud migration	<p>Recommended if you can tolerate some cluster downtime or have bursty workloads. See the following to know more about the direct to cloud migration paths:</p> <ul style="list-style-type: none"> <li>Nifi workloads If you are running NiFi workloads without Hive, Impala, HBase, or Kafka: Migrate to Cloudera Data Hub and use the Flow Management cluster template.</li> <li>Kafka workloads If you are running Kafka workloads without Hive, Impala, HBase, or NiFi: Migrate to Cloudera Data Hub and use the Streams Messaging cluster template.</li> <li>HBase workloads If you are running HBase workloads without Hive or Impala: Migrate to Cloudera Data Hub and use the Cloudera Operational Database cluster template.</li> </ul>

Migration and Upgrade Paths	Description
	<ul style="list-style-type: none"> <li>Hive or Impala workloads If you are running Hive or Impala workloads without HBase: Migrate to Cloudera Data Warehouse.</li> <li>Other workloads Migrate to Cloudera Data Hub and use the custom cluster template.</li> </ul>

### Cloudera Upgrade and Migration Paths to Cloudera for Cloudera DataFlow Functions components

You can upgrade or migrate Cloudera DataFlow Functions components to Cloudera in the following ways:

What do you want to migrate?	Do you want to migrate or upgrade?
Streaming workloads	<ul style="list-style-type: none"> <li>In-place upgrade from CDH</li> <li>In-place upgrade from HDP</li> <li>Migrating Streaming workloads from HDF to Cloudera Base on premises</li> </ul>
Flow Management workloads	<ul style="list-style-type: none"> <li>Follow the applicable upgrade or migration path in <a href="#">Upgrade and migration paths</a>.</li> </ul>
Workloads for deprecated components	<ul style="list-style-type: none"> <li>Migrate Flume workloads to NiFi for Cloudera. This workflow is in development.</li> <li>Migrate Storm workloads to Flink for Cloudera. This content is in development.</li> </ul>

### Cloudera Migration and Upgrade Paths for Cloudera Data Science Workbench (Cloudera AI)

You can upgrade or migrate Cloudera Data Science Workbench to Cloudera AI in the following ways:

What do you want to migrate?	Do you want to migrate or upgrade?
Cloudera Data Science Workbench with CDH/HDP to Cloudera Data Science Workbench with Cloudera Base on premises	For in-place upgrade follow the upgrade <a href="#">documentation</a> . For migration follow the <a href="#">documented</a> migration steps.
Cloudera Data Science Workbench to Cloudera AI-Cloudera on cloud	The workflow and tools to migrate from Cloudera Data Science Workbench to Cloudera AI in Cloudera on cloud are in development. To perform this migration now, contact your Cloud account or professional services representative.
Cloudera Data Science Workbench to Cloudera AI-Cloudera Base on premises	The workflow and tools to migrate from Cloudera Data Science Workbench to Cloudera AI in Cloudera Base on premises is available as technical preview. For more information, see the <a href="#">migration documentation</a> .

## Regressions During an Upgrade

This document outlines the missing backports that have led to regressions during an upgrade or fresh installation. The following sections detail the comprehensive list of fixes and improvements that were implemented in a previous

release but are now missing in the upgraded Cloudera version. Furthermore, it also explains any behavioral changes for individual components that have been observed after the upgrade.

## Upgrade regression in Cloudera Runtime 7.3.1.400 SP2

### Summary: Removal of non-standard case() function syntax in Hive

The non-standard case() function syntax has been removed from Hive.

#### Previous behavior:

Hive previously allowed a non-standard SQL syntax for the CASE expression using a function-like call:

```
select case(...) from ....
```

#### New behavior:

The non-standard case() function has been removed. You must use the official SQL standard CASE expression syntax. The standard syntax remains supported:

- Simple CASE:

```
SELECT CASE a WHEN b THEN c [WHEN d THEN e]* [ELSE f] END FR  
OM ...
```

- Searched CASE:

```
SELECT CASE WHEN a=b THEN c [WHEN a=d THEN e]* [ELSE f] END  
FROM ...
```

### Apache Jira: [HIVE-28792](#)

### Summary: Security-related HTTP headers and HSTS enforcement in Streams Replication Manager

#### Previous behavior:

Streams Replication Manager included the following security headers in API responses and enforced HTTP Strict Transport Security (HSTS) when SSL was enabled:

- Strict-Transport-Security: max-age=31536000; includeSubDomains - Enforces HTTPS communication for all requests, including subdomains, for a duration of one year. This header is only added when SSL is enabled.
- X-Frame-Options: SAMEORIGIN - Prevents clickjacking attacks by disallowing the page from being embedded in elements on other domains.
- X-XSS-Protection: 1; mode=block - Enables the XSS filter of the browser and blocks the page if an attack is detected.
- Content-Type: application/json - Ensures proper MIME type handling for responses.
- Cache-Control - Controls caching behavior.
  - Configured for API endpoints as no-store.
  - Configured for Swagger UI endpoints as max-age=7200.
- X-Content-Type-Options: nosniff - Prevents MIME type sniffing.
- Content-Security-Policy - Implements strict policies to control the sources of content that can be loaded.
  - Configured for API endpoints as default-src 'none';
  - Configured for Swagger UI endpoints as default-src 'self' 'unsafe-inline'; font-src 'self' data ::; img-src 'self' data:; style-src-attr 'self' 'unsafe-hashes'; style-src 'self' unsafe-inline'.

#### New behavior:

The headers are no longer included. Additionally, HTTP Strict Transport Security (HSTS) is not enforced when SSL is enabled.

**Summary: Update to transient\_lastDdlTime table property behavior for Iceberg tables****Previous behavior:**

Running the ALTER TABLE Impala statement on an Iceberg table automatically updated the transient\_lastDdlTime table property.

**New behavior:**

The ALTER TABLE statement no longer updates the transient\_lastDdlTime table property.

Iceberg does not persist the table name within its metadata files. Consequently, operations like renaming a table only modify the catalog entry, while the metadata (table state) stored in the file system remains unchanged.

**Apache Jira:** [IMPALA-11672](#)

**Summary: Changes to ALTER TABLE SET PARTITION SPEC field ID distribution****Previous behavior:**

The partition specification update mechanism was modified to utilize an updated Iceberg API that preserves existing partition specifications by generating new field IDs for modified partitioning terms. The change addressed issues with field ID distribution. For V1 Iceberg tables, the method now correctly adds the VOID partition transform for fields removed during updates. For V2 tables, updated fields are reused, but their order remains consistent with when they were first introduced in the specification.

**New behavior:**

Impala reuses existing partition field IDs when modifying partition transforms (for example, changing bucket counts or truncate widths) instead of assigning new partition field IDs for the transform.

**Apache Jira:** [IMPALA-12605](#)

**Summary: Inconsistent commons-lang3 versions in Avro dependencies****Previous behavior:**

All components correctly inherited the centralized commons-lang3 version.

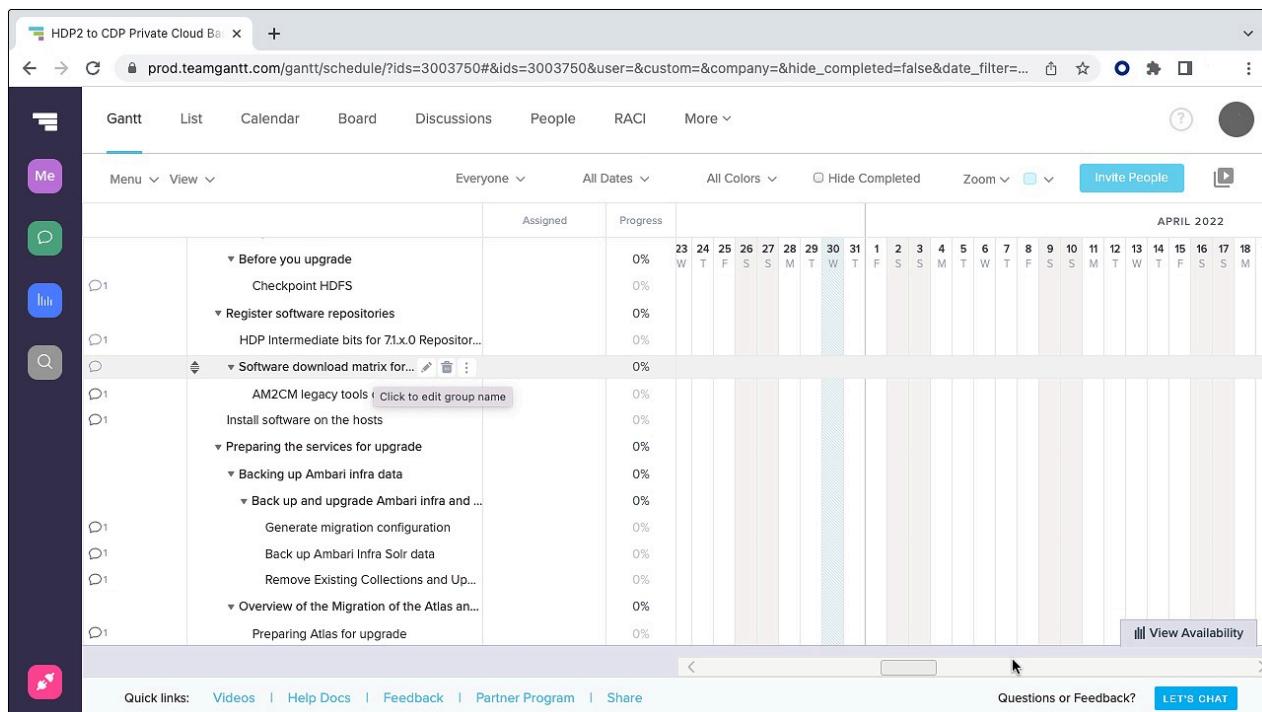
**New behavior:**

The centralized version was updated to 3.13, but the following components did not update and remain on divergent commons-lang3 versions:

- Apache Avro Compiler
- Apache Avro IPC
- Apache Avro Mapred API
- Apache Avro Maven Plugin
- Apache Avro Thrift Compatibility

## Downloadable Cloudera upgrade checklists

Upgrade checklists include all the upgrade steps in project format so you can plan and track your upgrade activities.



Use the following files to assist your Cloudera upgrade project in any way helpful. The file structure has rows reflecting topics from Cloudera upgrade documentation and columns labeling the topics, the topic hierarchy, and URLs back to this website. The columns particularly target TeamGantt.com. Upload one of the files to that site to see a graphical view of your upgrade project with expandable/collapsible hierarchical tasks, the ability to assign owners, check off completed items, insert dates, click to view Cloudera documentation, etc. Using TeamGantt.com, however, is not required.

- [From CDH 5 to CDP Private Cloud Base](#)
- [From CDH 6 to CDP Private Cloud Base](#)
- [From HDP 2 to CDP Private Cloud Base](#)
- [From HDP 3 to CDP Private Cloud Base](#)
- [From Cloudera Base on premises to Cloudera Base on premises](#)

## Supported in-place upgrade paths

Supported upgrade paths for Cloudera Base on premises, Cloudera Data Services on premises, CDH, and HDP.



### Important:

Before performing an upgrade of Cloudera Manager or Cloudera Runtime, creating a backup of all the metadata databases is important. This includes the Cloudera Manager database, and the various Cloudera Runtime component databases such as Hive Metadata Server, Ranger Admin, Ranger KMS, Schema Registry, and so on. The backups are necessary if there is a reason to rollback to the prior version.

The tables below detail the supported upgrade paths for upgrades to Cloudera Base on premises and Cloudera Data Services on premises. Before upgrading, ensure that you select a compatible version of Cloudera Manager. See [Cloudera Manager Support Matrix](#) on page 17.

For supported upgrades to CDH, see [Supported Upgrade Paths](#)

Upgrades from Cloudera Director are not supported.

**For Cloudera Base on premises**

**Table 1: Upgrade paths for Cloudera Base on premises**

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>Cloudera Base on premises 7.3.1.400 SP2</li> <li>Cloudera Base on premises 7.3.1.200 SP1</li> <li>Cloudera Base on premises 7.3.1</li> <li><a href="#">CDP Private Cloud Base 7.1.9 SP1 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.9 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.8 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.7 SP3 Latest cumulative hotfix</a></li> </ul>	<a href="#">Cloudera Base on premises 7.3.1.500 SP3 / Cloudera Manager 7.13.1.500 CHF5</a>	None
<ul style="list-style-type: none"> <li>Cloudera Base on premises 7.3.1.200 SP1</li> <li>Cloudera Base on premises 7.3.1</li> <li><a href="#">CDP Private Cloud Base 7.1.9 SP1 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.9 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.8 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.7 SP3 Latest cumulative hotfix</a></li> </ul>	<a href="#">Cloudera Base on premises 7.3.1.400 SP2 / Cloudera Manager 7.13.1.400 CHF4</a>	<ul style="list-style-type: none"> <li>If you want to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 to Cloudera Base on premises 7.3.1.400 SP2, you must not have any CHF installed on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3. The CHFs on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 are not in sync with Cloudera Base on premises 7.3.1.400 SP2.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 with any CHF to Cloudera Base on premises 7.3.1.400 SP2 could lead to feature regression or loss of few bug fixes.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.8 CHF 18 and higher to Cloudera Base on premises 7.3.1.400 SP2 could lead to feature regression or loss of few bug fixes.</li> </ul> <p> <b>Caution:</b> Cloudera Data Services on premises 1.5.5 is not supported on Cloudera Base on premises 7.3.1.400 SP2. You must not install or upgrade to Cloudera Base on premises 7.3.1.400 SP2 if you are using Cloudera Data Services on premises 1.5.5 on your cluster as it is incompatible.</p>

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>Cloudera Base on premises 7.3.1</li> <li><a href="#">CDP Private Cloud Base 7.1.9 SP1 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.9 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.8 Latest cumulative hotfix</a></li> <li><a href="#">CDP Private Cloud Base 7.1.7 SP3 Latest cumulative hotfix</a></li> </ul>	<a href="#">Cloudera Base on premises 7.3.1.200 SP1 / Cloudera Manager 7.13.1.200 CHF2</a>	<ul style="list-style-type: none"> <li>If you want to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 to Cloudera Base on premises 7.3.1.200 SP1, you must not have any CHF installed on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3. The CHFs on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 are not in sync with Cloudera Base on premises 7.3.1.200 SP1.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 with any CHF to Cloudera Base on premises 7.3.1.200 SP1 could lead to feature regression or loss of few bug fixes.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.8 CHF 18 and higher to Cloudera Base on premises 7.3.1.200 SP1 could lead to feature regression or loss of few bug fixes.</li> </ul> <p> <b>Caution:</b> Cloudera Data Services on premises 1.5.4 is not supported on Cloudera Base on premises 7.3.1.200 SP1. You must not install or upgrade to Cloudera Base on premises 7.3.1.200 SP1 if you are using Cloudera Data Services on premises 1.5.4 on your cluster as it is incompatible.</p>
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.9 SP1</li> <li>CDP Private Cloud Base 7.1.9</li> <li><a href="#">CDP Private Cloud Base 7.1.8 cumulative hotfix 17</a></li> <li>CDP Private Cloud Base 7.1.7 SP3</li> </ul>	<a href="#">Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1 / Cloudera Manager 7.13.1</a>	<ul style="list-style-type: none"> <li>If you want to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 to Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1, you must not have any CHF installed on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3. The CHFs on CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 are not in sync with Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.9 SP1 or 7.1.7 SP3 with any CHF to Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1 could lead to feature regression or loss of few bug fixes.</li> <li>Trying to upgrade from CDP Private Cloud Base 7.1.8 CHF 18 and higher to Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1 could lead to feature regression or loss of few bug fixes.</li> </ul> <p> <b>Caution:</b> Cloudera Data Services on premises 1.5.4 is not supported on Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1. You must not install or upgrade to Cloudera Base on premises 7.3.1 or 7.3.1.100 CHF1 if you are using Cloudera Data Services on premises 1.5.4 on your cluster as it is incompatible.</p>

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.8 CHF2 or higher / Cloudera Manager 7.7.1</li> <li>CDP Private Cloud Base 7.1.8 CHF2 or higher / Cloudera Manager 7.7.3</li> <li>CDP Private Cloud Base 7.1.7.2000 (SP2) / Cloudera Manager 7.6.7</li> <li>CDP Private Cloud Base 7.1.7.1000 (SP1) / Cloudera Manager 7.6.1</li> </ul>	CDP Private Cloud Base 7.1.9 / <a href="#">Cloudera Manager 7.11.3</a>	
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.8 / Cloudera Manager 7.7.1</li> <li>CDP Private Cloud Base 7.1.7.1000 (SP1) / Cloudera Manager 7.6.1</li> <li>CDP Private Cloud Base 7.1.7.78 (HOTFIX-4836 "Log4j")</li> <li>CDP Private Cloud Base 7.1.7 / Cloudera Manager 7.4.4</li> </ul>	<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.8 / <a href="#">Cloudera Manager 7.7.3</a></li> <li>Cloudera Manager 7.7.3-CHF1</li> </ul>	
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.7.78 / Cloudera Manager 7.5.4-20668437 (w/ Log4J fixes)</li> <li>CDP Private Cloud Base 7.1.7.1000 / Cloudera Manager 7.6.1 (SP1)</li> <li>CDP Private Cloud Base 7.1.7 / Cloudera Manager 7.4.4</li> <li>CDP Private Cloud Base 7.1.6 / Cloudera Manager 7.3.1</li> <li>CDP Private Cloud Base / Cloudera Manager 7.1.1, 7.1.2, 7.1.3, 7.1.4</li> <li>CDP Private Cloud Base 7.1.5 / Cloudera Manager 7.2.4</li> </ul>	CDP Private Cloud Base 7.1.8 / <a href="#">Cloudera Manager 7.7.1</a>	Upgrading Cloudera Manager to version 7.7.1 or higher from clusters where CDH 5.x is deployed is not supported. To upgrade such clusters: <ol style="list-style-type: none"> <li>Upgrade Cloudera Manager to version 6.3.4.</li> <li>Upgrade CDH to version 6.3.4</li> <li>Upgrade Cloudera Manager to version 7.7.1</li> </ol>
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.0.3 - 7.1.6</li> <li>Cloudera Manager 7.0.3-7.1.4, 7.2.4, 7.3.1</li> </ul>	CDP Private Cloud Base 7.1.7 / Cloudera Manager 7.4.4	
<ul style="list-style-type: none"> <li>Lower versions of CDP Private Cloud Base 7.0.3 - 7.1.6</li> <li>Lower versions of Cloudera Manager 7.0.3-7.1.4, 7.3.1</li> </ul>	CDP Private Cloud Base 7.1.1 - 7.1.6 / Cloudera Manager 7.1.1-7.1.4, 7.3.1	



**Note:** You can upgrade to Cloudera Runtime 7.1.7 Service Pack 1 (7.1.7.1000) from all of Cloudera Runtime 7.1.x, CDH5, and CDH6 versions. An upgrade to Cloudera Manager 7.6.1 is required for Service Pack 1 (7.1.7.1000).

**Table 2: Upgrade paths for Cloudera Base on premises Service Packs**

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.9 / Cloudera Manager 7.11.3</li> <li>CDP Private Cloud Base 7.1.7.3000 (SP3) / Cloudera Manager 7.11.3 CHF5</li> <li>CDP Private Cloud Base 7.1.8 <a href="#">Latest cumulative hotfix</a> / Cloudera Manager 7.7.1 <a href="#">Latest cumulative hotfix</a></li> <li>CDP Private Cloud Base 7.1.8 <a href="#">Latest cumulative hotfix</a> / Cloudera Manager 7.7.3 <a href="#">Latest cumulative hotfix</a></li> </ul>	CDP Private Cloud Base 7.1.9.1000 (SP1) / Cloudera Manager 7.11.3 <a href="#">Latest cumulative hotfix</a>	

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.7.2000 (SP2) / Cloudera Manager 7.6.7 <a href="#">Latest cumulative hotfix</a></li> <li>CDP Private Cloud Base 7.1.7.1000 (SP1) / Cloudera Manager 7.6.1</li> <li>CDP Private Cloud Base 7.1.7 / Cloudera Manager 7.4.4</li> <li>CDP Private Cloud Base 7.1.6 / Cloudera Manager 7.3.1</li> </ul>	CDP Private Cloud Base 7.1.7.3000 (SP3) / Cloudera Manager 7.11.3 <a href="#">Latest cumulative hotfix</a>	
<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.7.1000 (SP1) / Cloudera Manager 7.6.1</li> <li>CDP Private Cloud Base 7.1.7 (Log4J2) / Cloudera Manager 7.4.4 (Log4J)</li> <li>CDP Private Cloud Base 7.1.6 / Cloudera Manager 7.3.1</li> </ul>	CDP Private Cloud Base 7.1.7.2000 (SP2) <a href="#">Latest cumulative hotfix</a> / <a href="#">Cloudera Manager 7.6.7</a>	
CDP Private Cloud Base 7.1.7/Cloudera Manager 7.4.4	CDP Private Cloud Base 7.1.7.1000 (SP1) <a href="#">Latest cumulative hotfix</a> / <a href="#">Cloudera Manager 7.6.1</a>	<ul style="list-style-type: none"> <li>Parcel upgrade only required (running the Upgrade Wizard is not required). See <a href="#">Applying a Service Pack</a>.</li> <li>Cloudera Manager 7.6.5 is intended for use with CDP Private Cloud Base with CDP Private Cloud Data Services. Cloudera does not recommend using Cloudera Manager 7.6.5 in CDP Private Cloud Base environments without CDP Private Cloud Data Services installations.</li> </ul>

### For CDP Private Cloud Data Services

**Table 3: Upgrade paths for Cloudera Data Services on premises 1.5.5**

Source version			Target version		
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises	
<b>Fresh install of Cloudera Data Services on premises 1.5.5</b>					
NA	NA	NA	7.13.1 Cumulative hotfix 3	7.1.9 Service Pack 1 Cumulative hotfix 7	
NA	NA	NA	7.13.1 Cumulative hotfix 3	7.1.7 SP3 CHF 10	
<b>Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.5</b>					
Cloudera Manager 7.11.3 CHF6	7.1.9 CHF6	1.5.4	7.13.1 Cumulative hotfix 3	7.1.9 Service Pack 1 Cumulative hotfix 7	
Cloudera Manager 7.11.3 CHF6	7.1.8 CHF22	1.5.4	7.13.1 Cumulative hotfix 3	7.1.9 Service Pack 1 Cumulative hotfix 7	
Cloudera Manager 7.11.3 CHF6	7.1.7 SP3	1.5.4	7.13.1 Cumulative hotfix 3	7.1.7 SP3 CHF 10	
<b>Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premises</b>					
7.11.3 Cumulative hotfix 11	7.1.9 Service Pack 1 Cumulative hotfix 7	1.5.4 SP1	7.13.1 Cumulative hotfix 3	7.1.9 Service Pack 1 Cumulative hotfix 7	
7.13.1 Cumulative hotfix 1	7.1.7 SP3 CHF 10	1.5.4 SP2	7.13.1 Cumulative hotfix 3	7.1.7 SP3 CHF 10	
7.11.3 Cumulative hotfix 7 and 7.11.3 Cumulative hotfix 9.1	7.1.7 SP3 CHF 10	1.5.4 CHF1 and CHF3	7.13.1 Cumulative hotfix 3	7.1.7 SP3 CHF 10	

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
7.11.3 Cumulative hotfix 1	7.1.9 Service Pack 1 Cumulative hotfix 7	1.5.3 SP1 and 1.5.3	7.13.1 Cumulative hotfix 3	7.1.9 Service Pack 1 Cumulative hotfix 7

**Table 4: Upgrade paths for Cloudera Data Services on premises 1.5.4 SP2**

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
<b>Fresh install of Cloudera Data Services on premises 1.5.4 SP2</b>				
NA	NA	NA	7.13.1 Cumulative hotfix 1	7.1.9 Cumulative hotfix 7
NA	NA	NA	7.13.1 Cumulative hotfix 1	7.1.9 Service Pack 1
NA	NA	NA	7.13.1 Cumulative hotfix 1	7.1.7 SP3
<b>Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.4 SP2</b>				
7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6	1.5.4	7.13.1 Cumulative hotfix 1	7.1.9 Cumulative hotfix 7
7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher	1.5.4	7.13.1 Cumulative hotfix 1	7.1.9 Service Pack 1 Cumulative hotfix 4
7.11.3 Cumulative hotfix 6	7.1.7 SP3	1.5.4	7.13.1 Cumulative hotfix 1	7.1.7 SP3
7.11.3 Cumulative hotfix 4	7.1.7 SP2, 7.1.8 CHF 17, 7.1.9.4	1.5.3	7.13.1 Cumulative hotfix 1	7.1.7 SP 3, 7.1.9 Cumulative hotfix 7, or 7.1.9 SP1 CHF 4
<b>Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premises</b>				
7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 7	1.5.3	7.13.1 Cumulative hotfix 1	7.1.9 Cumulative hotfix 7
7.11.3 Cumulative hotfix 4	7.1.9 Service Pack 1	1.5.3	7.13.1 Cumulative hotfix 1	7.1.9 Service Pack 1
7.11.3 Cumulative hotfix 4	7.1.7 SP3	1.5.3	7.13.1 Cumulative hotfix 1	7.1.7 SP3
7.11.3 Cumulative hotfix 1	7.1.7 SP3	1.5.2	7.13.1 Cumulative hotfix 1	7.1.7 SP3

**Table 5: Upgrade paths for Cloudera Data Services on premises 1.5.4 SP1**

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
<b>Fresh install of Cloudera Data Services on premises 1.5.4 SP1</b>				
NA	NA	NA	7.11.3 Cumulative hotfix 11	7.1.9 Cumulative hotfix 7
NA	NA	NA	7.11.3 Cumulative hotfix 11	7.1.9 Service Pack 1 CHF 4
NA	NA	NA	7.11.3 Cumulative hotfix 11	7.1.7 SP3
<b>Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.4 SP1</b>				
7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6	1.5.4	7.11.3 Cumulative hotfix 11	7.1.9 Cumulative hotfix 7
7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher	1.5.4	7.11.3 Cumulative hotfix 11	7.1.9 Service Pack 1 CHF4
7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher and 7.1.7 SP3	1.5.4	7.11.3 Cumulative hotfix 11	7.1.7 SP3
7.11.3 Cumulative hotfix 4	7.1.9.4, 7.1.7 SP2, 7.1.8 CHF 17	1.5.3	7.11.3 Cumulative hotfix 11	7.1.7 SP3, 7.1.9 Cumulative hotfix 7, 7.1.9 Service Pack 1 CHF4
<b>Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premises</b>				

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 7	1.5.3	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 7
7.11.3 Cumulative hotfix 4	7.1.9 Service Pack 1	1.5.3	7.11.3 Cumulative hotfix 6	7.1.9 Service Pack 1
7.11.3 Cumulative hotfix 4	7.1.7 SP3	1.5.3	7.11.3 Cumulative hotfix 6	7.1.7 SP3
7.11.3 Cumulative hotfix 1	7.1.7 SP3	1.5.2	7.11.3 Cumulative hotfix 6	7.1.7 SP3

**Table 6: Upgrade paths for Cloudera Data Services on premises 1.5.4**

Source version			Target version	
Cloudera Manager	Cloudera Base on premises	Cloudera Data Services on premises	Cloudera Manager	Cloudera Base on premises
<b>Fresh install of Cloudera Data Services on premises 1.5.4</b>				
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
NA	NA	NA	7.11.3 Cumulative hotfix 6	7.1.7 SP3
<b>Using Cloudera Base on premises, but new to Cloudera Data Services on premises 1.5.4</b>				
7.11.3 Cumulative hotfix 4	7.1.9.4	1.5.3	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
7.7.3 <a href="#">Latest cumulative hotfix</a>	7.1.8 Cumulative hotfix 17	1.5.3	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
7.11.3 Cumulative hotfix 4	7.1.7 SP2	1.5.3	7.11.3 Cumulative hotfix 6	7.1.7 SP3
7.11.3 Cumulative hotfix 1	7.1.7 SP2, 7.1.8 Cumulative hotfix 11, 7.1.9	1.5.2	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6, 7.1.8 Cumulative hotfix 22 or higher, and 7.1.7 SP3
<b>Using Cloudera Data Services on premises and wanting to upgrade to the latest version without upgrading Cloudera Base on premises</b>				
7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 6	1.5.3	7.11.3 Cumulative hotfix 6	7.1.9 Cumulative hotfix 6
7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 22 or higher	1.5.3	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher
7.11.3 Cumulative hotfix 4	7.1.7 SP3	1.5.3	7.11.3 Cumulative hotfix 6	7.1.7 SP3
7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 22 or higher	1.5.2	7.11.3 Cumulative hotfix 6	7.1.8 Cumulative hotfix 22 or higher

**Table 7: Upgrade paths for CDP Private Cloud Data Services 1.5.3**

Source version			Target version	
Cloudera Manager	CDP Private Cloud Base	CDP Private Cloud Data Services	Cloudera Manager	CDP Private Cloud Base
<b>Greenfield deployment of CDP Private Cloud Data Services 1.5.3</b>				
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher
NA	NA	NA	7.11.3 Cumulative hotfix 4	7.1.7 SP2
<b>Using CDP Private Cloud Base, but new to CDP Private Cloud Data Services 1.5.3</b>				
7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 3	NA	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3

Source version			Target version		
Cloudera Manager	CDP Private Cloud Base	CDP Private Cloud Data Services	Cloudera Manager	CDP Private Cloud Base	
7.7.3 <a href="#">Latest cumulative hotfix</a>	7.1.8 Cumulative hotfix 19 or higher	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher	
7.7.1 <a href="#">Latest cumulative hotfix</a>	7.1.8 Cumulative hotfix 19 or higher	NA	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher	
7.6.7 <a href="#">Latest cumulative hotfix</a>	7.1.7 SP2	NA	7.11.3 Cumulative hotfix 4	7.1.7 SP2	
<b>Using CDP Private Cloud Data Services and wanting to upgrade to the latest version without upgrading CDP Private Cloud Base</b>					
7.11.3 Cumulative hotfix 1	7.1.9 Cumulative hotfix 3	1.5.2	7.11.3 Cumulative hotfix 4	7.1.9 Cumulative hotfix 3	
7.11.3 Cumulative hotfix 1	7.1.8 Cumulative hotfix 19 or higher	1.5.2	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher	
7.11.3 Cumulative hotfix 1	7.1.7 SP2	1.5.2	7.11.3 Cumulative hotfix 4	7.1.7 SP2	
7.10.1	7.1.8 Cumulative hotfix 19 or higher	1.5.1	7.11.3 Cumulative hotfix 4	7.1.8 Cumulative hotfix 19 or higher	
7.10.1	7.1.7 SP2	1.5.1	7.11.3 Cumulative hotfix 4	7.1.7 SP2	

### For Upgrade paths for CDH and HDP



**Note:** You can perform an In-place upgrade from CDH 6 to Cloudera Base on premises 7.3.1 in two separate steps. First, you must upgrade from CDH 6 to CDP Private Cloud Base 7.1.9 SP1 or [CDP Private Cloud Base 7.1.8 cumulative hotfix 17](#) or CDP Private Cloud Base 7.1.7 SP3 and then upgrade to Cloudera Base on premises 7.3.1.

**Table 8: Upgrade paths for CDH**

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>CDH 6.3.4 / Cloudera Manager 6.3.4</li> <li>CDH 6.2.1 / Cloudera Manager 6.2.1</li> </ul>	<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.9.1000 (SP1) / Cloudera Manager 7.11.3 <a href="#">Latest cumulative hotfix</a></li> </ul>	
<ul style="list-style-type: none"> <li>CDH 6.3.4 / Cloudera Manager 6.3.4</li> <li>CDH 6.2.1 / Cloudera Manager 6.2.1</li> </ul>	<ul style="list-style-type: none"> <li>CDP Private Cloud Base 7.1.7.3000 (SP3) / Cloudera Manager 7.11.3 <a href="#">Latest cumulative hotfix</a></li> <li>CDP Private Cloud Base 7.1.9 / Cloudera Manager 7.11.3</li> </ul>	
<ul style="list-style-type: none"> <li>CDH 6.3.4 / Cloudera Manager 6.3.4</li> <li>CDH 6.2.1 / Cloudera Manager 6.2.1</li> <li>CDH 6.1.1 / Cloudera Manager 6.1.1</li> <li>CDH 5.16.2 / Cloudera Manager 5.16.2</li> <li>CDH 5.16.2 / Cloudera Manager 6.3.1</li> </ul>	CDP Private Cloud Base 7.1.7.2000 (SP2) / <a href="#">Cloudera Manager 7.6.7</a>	
• CDH / Cloudera Manager 6.1 - 6.3	CDP Private Cloud Base 7.1.8 / <a href="#">Cloudera Manager 7.7.1</a>	<p>Upgrading Cloudera Manager to version 7.7.1 or higher from clusters where CDH 5.x is deployed is not supported. To upgrade such clusters:</p> <ol style="list-style-type: none"> <li>1. Upgrade Cloudera Manager to version 6.3.4.</li> <li>2. Upgrade CDH to version 6.3.4</li> <li>3. Upgrade Cloudera Manager to version 7.7.1</li> </ol>
<ul style="list-style-type: none"> <li>CDH / Cloudera Manager 5.13-5.16</li> <li>CDH / Cloudera Manager 6.1-6.3</li> </ul>	CDP Private Cloud Base 7.1.7 /Cloudera Manager 7.4.4	

Upgrade supported from:	Upgrade to:	Notes
<ul style="list-style-type: none"> <li>CDH / Cloudera Manager 5.13 - 5.16</li> <li>CDH / Cloudera Manager 6.1, 6.2</li> </ul>	CDH 6.3	
CDH / Cloudera Manager 5.0 - 5.12	CDH / Cloudera Manager 5.13 - 5.16	Upgrades from Cloudera Manager/CDH 5.0 - 5.12 to CDP Private Cloud Base require that you first upgrade to Cloudera Manager/CDH 5.13 or higher.



**Note:** You can perform an In-place upgrade from HDP 3 to Cloudera Base on premises 7.3.1 in two separate steps. First, you must upgrade from HDP 3 to CDP Private Cloud Base 7.1.9 SP1 or [CDP Private Cloud Base 7.1.8 cumulative hotfix 17](#) or CDP Private Cloud Base 7.1.7 SP3 and then upgrade to Cloudera Base on premises 7.3.1.

**Table 9: Upgrade paths for HDP**

Upgrade supported from:	Upgrade to:	Notes
HDP 3.1.5	CDP Private Cloud Base 7.1.7 SP1, CDP Private Cloud Base 7.1.8, CDP Private Cloud Base 7.1.7 SP2, 7.1.9, CDP Private Cloud Base 7.1.7 SP3, and CDP Private Cloud Base 7.1.9 SP1.	One stage upgrade.
HDP 2.6.5	CDP Private Cloud Base 7.1.8, CDP Private Cloud Base 7.1.7 SP2, and CDP Private Cloud Base 7.1.7 SP3.	One stage upgrade.
HDP 3.1.5	CDP Private Cloud Base 7.1.8 / <a href="#">Cloudera Manager 7.7.1</a>	Two stage upgrade from Ambari required.
HDP 2.6.5	CDP Private Cloud Base 7.1.8 / <a href="#">Cloudera Manager 7.7.1</a> . Requires an interim upgrade to CDP Private Cloud Base 7.1.7/Cloudera Manager 7.4.4.	Two stage upgrade from Ambari required.
HDP 2.6.5	CDP Private Cloud Base 7.1.1 - 7.1.7	You can upgrade from HDP 2.6.5 and Ambari 2.6.2.x to CDP Private Cloud Base 7.1.x. This upgrade requires several major steps, including upgrading to an interim version of Ambari. After the upgrade, your cluster will be managed by Cloudera Manager and the components will be upgraded to Cloudera Runtime 7.1.1 or higher.

## Cloudera Manager Support Matrix

This support matrix lists the Cloudera Manager version support for Cloudera Base on premises and the Cloudera Data Services on premises.



**Note:** Not all combinations of Cloudera Manager, Cloudera Runtime, and Cloudera Data Services on premises are supported. Ensure that the version of Cloudera Manager you are using supports the version of Cloudera Runtime and Cloudera Data Services on premises you have selected. For more information, see the [Cloudera Support Matrix](#).

The versions of Cloudera Runtime, Cloudera Data Services on premises, and CDH clusters that can be managed by Cloudera Manager are limited to the following:

**For CDP Private Cloud Base**

**Table 10: Cloudera Manager support for Cloudera Base on premises**

Cloudera Manager Version	Supported CDH/Cloudera Runtime versions
Cloudera Manager 7.13.1 Latest cumulative hotfix	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.3.1.500 SP3</li> <li>• Cloudera Runtime 7.3.1.400 SP2</li> <li>• Cloudera Runtime 7.3.1.200 SP1</li> <li>• Cloudera Runtime 7.3.1</li> <li>• Cloudera Runtime 7.1.9 SP1</li> <li>• Cloudera Runtime 7.1.9</li> <li>• Cloudera Runtime 7.1.8</li> <li>• Cloudera Runtime 7.1.7 SP3</li> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> </ul>
Cloudera Manager 7.13.1	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.3.1</li> <li>• Cloudera Runtime 7.1.9 SP1</li> <li>• Cloudera Runtime 7.1.9</li> <li>• Cloudera Runtime 7.1.8</li> <li>• Cloudera Runtime 7.1.7 SP3</li> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> <li>• CDH 6.3</li> <li>• CDH 6.2</li> <li>• CDH 6.1</li> <li>• CDH 6.0</li> </ul>
Cloudera Manager 7.11.3 Latest cumulative hotfix	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.9 SP1</li> <li>• Cloudera Runtime 7.1.7 SP3</li> <li>• Cloudera Runtime 7.1.9</li> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.8</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> <li>• Cloudera Runtime 7.1.6</li> <li>• Cloudera Runtime 7.1.5</li> <li>• Cloudera Runtime 7.1.4</li> <li>• Cloudera Runtime 7.1.3</li> <li>• Cloudera Runtime 7.1.2</li> <li>• Cloudera Runtime 7.1.1</li> <li>• Cloudera Runtime 7.0.3</li> <li>• CDH 6.3</li> <li>• CDH 6.2</li> <li>• CDH 6.1</li> <li>• CDH 6.0</li> </ul>

Cloudera Manager Version	Supported CDH/Cloudera Runtime versions
<b>Cloudera Manager 7.11.3</b>  <p><b>Note:</b> You must install Python 3.8 (or 3.9 for RHEL 9.1) on all hosts before installing or upgrading to Cloudera Manager 7.11.3. For more information, see the <a href="#">Installing Python 3</a>.</p>	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.9</li> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.8</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> <li>• Cloudera Runtime 7.1.6</li> <li>• Cloudera Runtime 7.1.5</li> <li>• Cloudera Runtime 7.1.4</li> <li>• Cloudera Runtime 7.1.3</li> <li>• Cloudera Runtime 7.1.2</li> <li>• Cloudera Runtime 7.1.1</li> <li>• Cloudera Runtime 7.0.3</li> <li>• CDH 6.3</li> <li>• CDH 6.2</li> <li>• CDH 6.1</li> <li>• CDH 6.0</li> </ul>
<b>Cloudera Manager 7.7.3</b> should only be used when you need to use Python 3.8 for the Cloudera Manager agents. You must install Python 3.8 on all hosts before installing or upgrading to Cloudera Manager 7.7.3. Cloudera Manager 7.7.3-CHF2 supports only RHEL 8.4, RHEL 8.6, and RHEL 7.9. See the <a href="#">CDP Private Cloud Base Installation Guide</a> for more information.	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.8</li> </ul>
<b>Cloudera Manager 7.7.1</b>  <p><b>Note:</b> Cloudera recommends you to use latest cumulative hotfix of Cloudera Manager 7.7.1 with Cloudera Runtime 7.1.7-SP2.</p>	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.8</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> <li>• Cloudera Runtime 7.1.6</li> <li>• Cloudera Runtime 7.1.5</li> <li>• Cloudera Runtime 7.1.4</li> <li>• Cloudera Runtime 7.1.3</li> <li>• Cloudera Runtime 7.1.2</li> <li>• Cloudera Runtime 7.1.1</li> <li>• Cloudera Runtime 7.0.3</li> <li>• CDH 6.3</li> <li>• CDH 6.2</li> <li>• CDH 6.1</li> <li>• CDH 6.0</li> </ul>
<b>7.6.7</b>  <p><b>Important:</b> Do not upgrade to Cloudera Manager 7.6.7 if you are running <a href="#">Cloudera Data Services on premises</a> in your deployment.</p>	<ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.7 SP2</li> <li>• Cloudera Runtime 7.1.7 SP1</li> <li>• Cloudera Runtime 7.1.7</li> <li>• Cloudera Runtime 7.1.6</li> <li>• Cloudera Runtime 7.1.5</li> <li>• Cloudera Runtime 7.1.4</li> <li>• Cloudera Runtime 7.1.3</li> <li>• Cloudera Runtime 7.1.2</li> <li>• Cloudera Runtime 7.1.1</li> <li>• Cloudera Runtime 7.0.3</li> <li>• CDH 6.3</li> <li>• CDH 6.2</li> <li>• CDH 6.1</li> <li>• CDH 6.0</li> <li>• CDH 5.16.2</li> </ul>

Cloudera Manager Version	Supported CDH/Cloudera Runtime versions
7.6.1   <b>Important:</b> Do not upgrade to Cloudera Manager 7.6.1 if you are running <a href="#">Cloudera Data Services on premises</a> in your deployment.	<ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.7 SP1</li> <li>Cloudera Runtime 7.1.7</li> <li>Cloudera Runtime 7.1.6</li> <li>Cloudera Runtime 7.1.5</li> <li>Cloudera Runtime 7.1.4</li> <li>Cloudera Runtime 7.1.3</li> <li>Cloudera Runtime 7.1.2</li> <li>Cloudera Runtime 7.1.1</li> <li>Cloudera Runtime 7.0.3</li> <li>CDH 6.3</li> <li>CDH 6.2</li> <li>CDH 6.1</li> <li>CDH 6.0</li> <li>CDH 5.13 - 5.16</li> </ul>

## For CDP Private Cloud Data Services

**Table 11: Cloudera Manager support for Cloudera Data Services on premises**

Supported Cloudera Manager versions with Cloudera Data Services on premises	Supported Cloudera Runtime versions with Cloudera Data Services on premises
<a href="#">7.13.1 Cumulative hotfix 5 (7.13.1.501)</a>	Cloudera Data Services on premises <a href="#">1.5.5 SP1</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.3.1.500 SP3</li> <li>Cloudera Runtime 7.1.9 SP1 CHF11 or higher</li> <li>Cloudera Runtime 7.1.7 SP3 CHF10 or higher</li> </ul>
<a href="#">7.13.1 Cumulative hotfix 4 (7.13.1.402)</a>	Cloudera Data Services on premises <a href="#">1.5.5 CHF1</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.3.1.400 SP2</li> <li>Cloudera Runtime 7.1.9 SP1 CHF7 or higher</li> <li>Cloudera Runtime 7.1.7 SP3 CHF10 or higher</li> </ul>
<a href="#">7.13.1 Cumulative hotfix 3 (7.13.1.303)</a>	Cloudera Data Services on premises <a href="#">1.5.5</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.9 SP1 CHF7 or higher</li> <li>Cloudera Runtime 7.1.7 SP3 CHF10 or higher</li> </ul>
<a href="#">7.13.1 Cumulative hotfix 1 (7.13.1.100)</a>	Cloudera Data Services on premises <a href="#">1.5.4 SP2</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.9 SP1 CHF4 or higher</li> <li>Cloudera Runtime 7.1.9 CHF7 or higher</li> <li>Cloudera Runtime 7.1.7 SP3</li> </ul>
<a href="#">7.11.3 cumulative hotfix 11</a>	Cloudera Data Services on premises <a href="#">1.5.4 SP1</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.9 SP1 CHF4 or higher</li> <li>Cloudera Runtime 7.1.9 CHF7 or higher</li> <li>Cloudera Runtime 7.1.7 SP3</li> </ul>
<a href="#">7.11.3 cumulative hotfix 9.1</a> (version: 7.11.3.24)	Cloudera Private Cloud Data Services <a href="#">1.5.4 CHF3</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.9 SP1</li> <li>Cloudera Runtime 7.1.9 CHF6 or higher</li> <li>Cloudera Runtime 7.1.7 SP3</li> </ul>
<a href="#">7.11.3 cumulative hotfix 7 Data Services</a> (version: 7.11.3.14)	Cloudera Private Cloud Data Services <a href="#">1.5.4 CHF1</a> supported only with: <ul style="list-style-type: none"> <li>Cloudera Runtime 7.1.9 SP1</li> <li>Cloudera Runtime 7.1.9 CHF5 or higher</li> <li>Cloudera Runtime 7.1.7 SP3</li> </ul>

Supported Cloudera Manager versions with Cloudera Data Services on premises	Supported Cloudera Runtime versions with Cloudera Data Services on premises
7.11.3 cumulative hotfix 6	<p>Cloudera Private Cloud Data Services <a href="#">1.5.4</a> supported only with:</p> <ul style="list-style-type: none"> <li>• Cloudera Runtime 7.1.9 CHF6</li> <li>• Cloudera Runtime 7.1.8 CHF22 or higher</li> <li>• Cloudera Runtime 7.1.7 SP3</li> </ul>

## Cloudera Base on premises requirements and supported versions

Refer to the following topics for information about hardware, operating system, and database requirements, as well as product compatibility matrices.



### Important:

Before performing an upgrade of Cloudera Manager or the Cloudera Runtime, creating a backup of all the metadata databases is important. This includes the Cloudera Manager database, and the various Cloudera Runtime component databases such as Hive Metadata Server, Ranger Admin, Ranger KMS, Schema Registry, and so on. The backups are necessary if there is a reason to rollback to the prior version.

## Hardware Requirements

This topic specifies the hardware requirements for Cloudera Base on premises.

As you create the architecture of your cluster, you will need to allocate Cloudera Manager and Cloudera Runtime roles among the hosts in the cluster to maximize your use of resources. Cloudera provides some guidelines about how to assign roles to cluster hosts. See [Recommended Cluster Hosts and Role Distribution](#). When multiple roles are assigned to hosts, add together the total resource requirements (memory, CPUs, disk) for each role on a host to determine the required hardware.



**Attention:** All recommendations for the number of cores refer to logical cores, not physical cores.

For more information about sizing for a particular component, see the following minimum requirements:

### Cloudera Manager

Hardware requirements for Cloudera Manager Server and related components.

#### Cloudera Manager Server

**Table 12: Cloudera Manager Server Storage Requirements**

Component	Storage	Notes
Partition hosting /usr	1 GB	
Partition hosting /var	100 GB to 5 TB	Scales according to number of nodes managed. See table below.
Partition hosting /opt	100 GB minimum	Usage grows as the number of parcels downloaded increases. Budget 8 GB for each additional CDH parcel, and 1 GB for each additional non-CDH parcel.

Component	Storage	Notes
Cloudera Manager Database Server	<ul style="list-style-type: none"> <li>&lt; 500 hosts: 5 GB</li> <li>&gt; 500 hosts: 10 GB</li> </ul>	<p>Minimum memory and processor requirements should allow support for the following number of parallel database connections:</p> <ul style="list-style-type: none"> <li>&lt; 500 hosts: 100 database connections</li> <li>&gt; 500 hosts: 250 database connections</li> </ul>
Reports Manager Database Server	Minimum 1 GB	Reports Manager growth depends on number of HDFS users and monitored directories.

**Table 13: Host Based Cloudera Manager Server Requirements**

Number of Cluster Hosts	Database Host Configuration and HMON+SMON host sharing	Cloudera Manager Server Heap Size	Logical Processors	Cloudera Manager Server /var Directory	SMON and HMON / var Directory
Very small (#10)	Shared	8 GB	4	50 GB	50 GB
Small (#20)	Shared	10 GB	6	100 GB	100 GB
Medium (#200)	Dedicated	16 GB	8	1 TB	1 TB
Large (#500)	Dedicated	32 GB	12	2.5 TB	2.5 TB
Extra Large (>500)	Dedicated	48 GB	16	> 2.5 TB	> 2.5 TB



**Important:** For medium and larger clusters, Host Monitor (HMON) and Service Monitor (SMON) should run on a host that is separate from Cloudera Manager. For medium and larger clusters, the SQL database should not be shared between Cloudera Manager and CDH component services. Host Monitor and Service Monitor do not use SQL database. They use an on-disk LevelDB database in the /var partition.



**Note:** To increase the Cloudera Manager Server heap size you must update the parameter under /etc/default/cloudera-scm-server export CMF\_JAVA\_OPTS="-Xmx4G. Later restart the Cloudera Manager server for the changes to take effect.

### Service Monitor Requirements

The requirements for the Service Monitor are based on the number of monitored entities. To see the number of monitored entities, perform the following steps:

1. Open the Cloudera Admin Console and click Clusters .
2. Find the Cloudera Management Service Monitored Entities chart. If the chart does not exist, add it from the Chart Library.

For more information about Cloudera Manager entities, see *Cloudera Manager Entity Types*.



**Note:** Java Heap Size values (see the tables below) are rough estimates and some tuning might be necessary. From Cloudera Manager, Cloudera recommends using G1 garbage collector (G1GC) for Service Monitor. G1GC eliminates long JVM pauses, but uses a bit more CPU and RAM. It is the default for new installations. See [Tuning JVM Garbage Collection](#).



**Important:** Service Monitor is not supported when installed on the BTRFS filesystem.

**Table 14: Clusters with HDFS, YARN, or Impala**

Use the recommendations in this table for clusters where the only services with worker roles are HDFS, YARN, or Impala.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-2,000	0-100	1 GB	6 GB
2,000-4,000	100-200	1.5 GB	6 GB
4,000-8,000	200-400	1.5 GB	12 GB
8,000-16,000	400-800	2.5 GB	12 GB
16,000-20,000	800-1,000	3.5 GB	12 GB

**Table 15: Clusters with HBase, Solr, Kafka, or Kudu**

Use these recommendations when services such as HBase, Solr, Kafka, or Kudu are deployed in the cluster. These services typically have larger quantities of monitored entities.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-30,000	0-100	2 GB	12 GB
30,000-60,000	100-200	3 GB	12 GB
60,000-120,000	200-400	3.5 GB	12 GB
120,000-240,000	400-800	8 GB	20 GB

## Related Information

[Host Monitor and Service Monitor Memory Configuration](#)

### Host Monitor

The requirements for the Host Monitor are based on the number of monitored entities.

To see the number of monitored entities, perform the following steps:

1. Open the Cloudera Admin Console and click [Clusters](#).
2. Find the Cloudera Management Service Monitored Entities chart. If the chart does not exist, add it from the Chart Library.

For more information about Cloudera Manager entities, see [Cloudera Manager Entity Types](#).

**Important:** Host Monitor is not supported when installed on the BTRFS filesystem.



Number of Hosts	Number of Monitored Entities	Heap Size	Non-Java Heap Size
0-200	<6k	1 GB	2 GB
200-800	6k-24k	2 GB	6 GB
800-1000	24k-30k	3 GB	6 GB

Ensure that you have at least 25 GB of disk space available for the Host Monitor, Service Monitor, Reports Manager, and Events Server databases.

## Related Information

[Cloudera Manager Entity Types](#)

[Host Monitor and Service Monitor Memory Configuration](#)

### Reports Manager

The Reports Manager fetches the fsimage from the NameNode at regular intervals. It reads the fsimage and creates a Lucene index for it. To improve the indexing performance, Cloudera recommends provisioning a host as powerful as possible and dedicating an SSD disk to the Reports Manager.

**Table 16: Reports Manager**

Component	Java Heap	CPU	Disk
Reports Manager	3-4 times the size of the fsmimage.	<ul style="list-style-type: none"> <li>Minimum: 8 cores</li> <li>Recommended: 16 cores (32 cores, with hyperthreading enabled.)</li> </ul>	1 dedicated disk that is at least 20 times the size of the fsmimage. Cloudera recommends using SSD disks.

**Agent Hosts**

An unpacked parcel requires approximately three times the space of the packed parcel that is stored on the Cloudera Manager Server.

Component	Storage	Notes
Partition hosting /opt	30 GB minimum	Usage grows as new parcels are downloaded to cluster hosts.
/var/log	2 GB per role	Each role running on the host will need at least 2 GB of disk space.

**Event Server**

The following table lists the minimum requirements for the Event Server:

CPU	RAM	Storage
1 core	256 MB	<ul style="list-style-type: none"> <li>5 GB for the Event Database</li> <li>20 GB for the Event Server Index Directory. The location of this directory is set by the Event Server Index Directory Event Server configuration property.</li> </ul>

**Alert Publisher**

The following table lists the minimum requirements for the Alert Publisher:

CPU	RAM	Storage
1 core	1 GB	Minimum of 1 disk for log files

**Cloudera Runtime**

Hardware requirements for Cloudera Runtime components.

**Atlas**

You must verify the supported hardware components for Atlas.

Memory	CPU	Disk	Additional Dependencies
Small: 4 GB	Minimum: 4	No special requirement because HBase is used for storage.	Solr Shards: 4
Large: 32 GB	Medium: 8 Large: 16		(property: atlas_solr_shards) The shards for Atlas collections within Solr is determined by this number.

**HDFS heap sizing**

You can provision an HDFS cluster for optimal performance based on the desired storage capacity.

Component	Memory	CPU	Disk
JournalNode	1 GB (default) Set this value using the Java Heap Size of JournalNode in Bytes HDFS configuration property.	1 core minimum	1 dedicated disk
NameNode	<ul style="list-style-type: none"> <li>Minimum: 1 GB (for proof-of-concept deployments)</li> <li>Add an additional 1 GB for each additional 1,000,000 blocks</li> </ul> <p>Snapshots and encryption can increase the required heap memory.</p> <p>See <i>Sizing NameNode Heap Memory</i>.</p> <p>Set this value using the Java Heap Size of NameNode in Bytes HDFS configuration property.</p>	Minimum of 4 dedicated cores; more may be required for larger clusters	<ul style="list-style-type: none"> <li>Minimum of 2 dedicated disks for metadata</li> <li>1 dedicated disk for log files (This disk may be shared with the operating system.)</li> <li>Maximum disks: 4</li> </ul>
DataNode	<p>Minimum: 4 GB Maximum: 8 GB</p> <p>Increase the memory for higher replica counts or a higher number of blocks per DataNode. When increasing the memory, Cloudera recommends an additional 1 GB of memory for every 1 million replicas above 4 million on the DataNodes. For example, 5 million replicas require 5 GB of memory.</p> <p>Set this value using the Java Heap Size of DataNode in Bytes HDFS configuration property.</p>	Minimum: 4 cores. Add more cores for highly active clusters.	<p>Minimum: 4 Maximum: 24</p> <p>The maximum acceptable size will vary depending upon how large average block size is. The DN's scalability limits are mostly a function of the number of replicas per DN, not the overall number of bytes stored. That said, having ultra-dense DNs will affect recovery times in the event of machine or rack failure. Cloudera does not support exceeding 100 TB per data node. You could use 12 x 8 TB spindles or 24 x 4TB spindles. Cloudera does not support drives larger than 8 TB.</p> <p>Configure the disks in JBOD mode. Do not use RAID/LVM/ZFS.</p>



**Warning:** Running Cloudera Runtime on storage platforms other than direct-attached physical disks can provide suboptimal performance. Cloudera Enterprise and the majority of the Hadoop platform are optimized to provide high performance by distributing work across a cluster that can utilize data locality and fast local I/O.

## HBase

Know the hardware resource requirements for HBase server in Cloudera Base on premises.

Component	Java Heap	CPU	Disk
Master	<ul style="list-style-type: none"> <li>100-10,000 regions: 4 GB</li> <li>10,000 or more regions with 200 or more Region Servers: 8 GB</li> <li>10,000 or more regions with 300 or more Region Servers: 12 GB</li> </ul> <p>Set this value using the Java Heap Size of HBase Master in Bytes HBase configuration property.</p>	Minimum 4 dedicated cores. You can add more cores for larger clusters, when using replication, or for bulk loads.	1 disk for local logs, which can be shared with the operating system and/or other Hadoop logs

Component	Java Heap	CPU	Disk
Region Server	<ul style="list-style-type: none"> <li>Minimum: 8 GB</li> <li>Medium-scale production: 16 GB</li> <li>Heap larger than 16 GB requires special Garbage Collection tuning. See <i>Configuring the HBase BlockCache</i>.</li> </ul> <p>Set this value using the Java Heap Size of HBase RegionServer in Bytes HBase configuration property.</p>	Minimum: 4 dedicated cores	<ul style="list-style-type: none"> <li>4 or more spindles for each HDFS DataNode</li> <li>1 disk for local logs (this disk can be shared with the operating system and/or other Hadoop logs)</li> </ul>
Thrift Server	<p>1 GB - 4 GB</p> <p>Set this value using the Java Heap Size of HBase Thrift Server in Bytes HBase configuration property.</p>	Minimum 2 dedicated cores.	1 disk for local logs, which can be shared with the operating system and other Hadoop logs.



**Note:** Consider adding more HBase Thrift Servers for production environments and deployments with a large number of Thrift client to scale horizontally.

## Related Information

### [Configuring HBase BlockCache](#)

## Hive

Understand the Java Heap memory, CPU, and storage requirements needed to run HiveServer (HS2) and Hive Metastore (HMS).

Component	Java Heap		CPU	Disk	
HiveServer 2	Single Connection	4 GB	Minimum 4 dedicated cores	Minimum 1 disk	
	2-10 connections	4-6 GB		This disk is required for the following:	
	11-20 connections	6-12 GB		<ul style="list-style-type: none"> <li>HiveServer2 log files</li> <li>stdout and stderr output files</li> <li>Configuration files</li> <li>Operation logs stored in the operation_logs_dir directory, which is configurable</li> <li>Any temporary files that might be created by local map tasks under the /tmp directory</li> </ul>	
	21-40 connections	12-16 GB			
	41 to 80 connections	16-24 GB			
	Cloudera recommends splitting HiveServer2 into multiple instances and load balancing them once you start allocating more than 16 GB to HiveServer2. The objective is to adjust the size to reduce the impact of Java garbage collection on active processing by the service.				
	Set this value using the Java Heap Size of HiveServer2 in Bytes Hive configuration property.				

Component	Java Heap		CPU	Disk
Hive Metastore	Single Connection	4 GB	Minimum 4 dedicated cores	Minimum 1 disk
	2-10 connections	4-10 GB		This disk is required so that the Hive metastore can store the following artifacts:
	11-20 connections	10-12 GB		<ul style="list-style-type: none"> <li>Logs</li> <li>Configuration files</li> <li>Backend database that is used to store metadata if the database server is also hosted on the same node</li> </ul>
	21-40 connections	12-16 GB		
	41 to 80 connections	16-24 GB		
	Set this value using the Java Heap Size of Hive Metastore Server in Bytes Hive configuration property.			
Beeline CLI	Minimum: 2 GB		N/A	N/A

## Hue

Review the memory, CPU, and storage requirements needed for running a Hue server in Cloudera Base on premises.

Component	Memory	CPU	Disk
Hue Server	<ul style="list-style-type: none"> <li>Minimum: 4 GB</li> <li>Maximum 10 GB</li> <li>If the cluster uses the Hue load balancer, add additional memory</li> </ul>	Minimum: 1 Core to run Django When Hue is configured for high availability, add additional cores	Minimum: 10 GB for the database, which grows proportionally according to the cluster size and workloads. When Hue is configured for high availability, add space is required for the /tmp (temporary) directory, approximately 5GB.

The term "cluster size" refers to the number of nodes in the cluster. "Workload" in Hue means the number of queries run and the number of concurrent unique users using the application in a given period of time.

A minimum of 10GB is needed for the database. The Hive MetaStore service largely uses the database. The database grows in size quickly because of the query history that it retains. To optimize performance, you must regularly cleanup old documents and queries.



**Note:** Hue is limited by cgroup settings. In Cloudera Manager, all memory soft/hard limits are set to -1.

## Related Information

[Adding a Load Balancer for Hue](#)

## Impala

Sizing requirements for Impala can vary significantly depending on the size and types of workloads using Impala.

Component	Native Memory	JVM Heap	CPU	Disk
Impala Daemon	Set this value using the Impala Daemon Memory Limit configuration property. <ul style="list-style-type: none"> <li>Minimum: 32 GB</li> <li>Recommended: 128 GB</li> </ul>	Set this value using the Java Heap Size of Impala Daemon in Bytes configuration property for the Coordinator Impala Daemons. <ul style="list-style-type: none"> <li>Minimum: 4 GB</li> <li>Recommended: 8 GB</li> </ul>	<ul style="list-style-type: none"> <li>Minimum: 4</li> <li>Recommended: 16 or more</li> </ul> CPU instruction set: AVX2	<ul style="list-style-type: none"> <li>Minimum: 1 disk</li> <li>Recommended: 8 or more</li> </ul>

Component	Native Memory	JVM Heap	CPU	Disk
Catalog Server	<p>Set this value using the Java Heap Size of Catalog Server in Bytes configuration property.</p> <ul style="list-style-type: none"> <li>• Minimum: 4 GB</li> <li>• Recommended: 8 GB</li> </ul>		<ul style="list-style-type: none"> <li>• Minimum: 4</li> <li>• Recommended: 16 or more</li> </ul> <p>CPU instruction set: AVX2</p>	<ul style="list-style-type: none"> <li>• Minimum and Recommended: 1 disk</li> </ul>

For the networking topology for multi-rack cluster, Leaf-Spine is recommended for the optimal performance.

## Kafka

Kafka requires a fairly small amount of resources, especially with some configuration tuning. By default, Kafka, can run on as little as 1 core and 1GB memory with storage scaled based on requirements for data retention.

CPU is rarely a bottleneck because Kafka is I/O heavy, but a moderately-sized CPU with enough threads is still important to handle concurrent connections and background tasks.

Kafka brokers tend to have a similar hardware profile to HDFS data nodes. How you build them depends on what is important for your Kafka use cases.

Use the following guidelines:

To affect performance of these features:	Adjust these parameters:
Message Retention	Disk size
Client Throughput (Producer & Consumer)	Network capacity
Producer throughput	Disk I/O
Consumer throughput	Memory

A common choice for a Kafka node is as follows:

Component	Memory/Java Heap	CPU	Disk
Broker	<ul style="list-style-type: none"> <li>• RAM: 64 GB</li> <li>• Recommended Java heap: 4 GB</li> </ul> <p>Set this value using the Java Heap Size of Broker Kafka configuration property.</p>	12- 24 cores	<ul style="list-style-type: none"> <li>• 1 HDD For operating system</li> <li>• 1 HDD for Zookeeper dataLogDir</li> <li>• 10- HDDs, using Raid 10, for Kafka data</li> </ul>
Cruise Control	1 GB	1 core  <p><b>Note:</b> A moderately-sized CPU with enough threads is important to handle metric fetching from Kafka and background tasks.</p>	Because Cruise Control stores its data in Kafka the storage requirements will depend on the retention settings of the related Kafka topics.
Kafka Connect	0.5 - 4 GB heap size depending on the Connectors in use.	4 cores  <p><b>Note:</b> Depends on the Connectors in use.</p>	
MirrorMaker	1 GB heap Set this value using the Java Heap Size of MirrorMaker Kafka configuration property.	1 core per 3-4 streams	No disk space needed on MirrorMaker instance. Destination brokers should have sufficient disk space to store the topics being copied over.

Component	Memory/Java Heap	CPU	Disk
Schema Registry	1 GB heap	2 cores	1 MB  Serialization JAR files may be uploaded and may be of any size. The disk usage depends on the JAR files uploaded. The files may be stored locally on the same host where SchemaRegistry is running or in HDFS if available.
Streams Messaging Manager   <b>Note:</b> The hardware requirements for SMM depends on the number of Kafka partitions.	8 GB heap	8 cores	5 GB
Streams Replication Manager	<ul style="list-style-type: none"> <li>1 GB heap for SRM driver</li> <li>1 GB heap for SRM Service</li> </ul>	The performance of the SRM driver is mostly impacted by network throughput and latency.	No resources required

Networking requirements: Gigabit Ethernet or 10 Gigabit Ethernet. Avoid clusters that span multiple data centers.

Kafka and Zookeeper: It is common to run ZooKeeper on 3 broker nodes that are dedicated for Kafka. However, for optimal performance Cloudera recommends the usage of dedicated Zookeeper hosts. This is especially true for larger, production environments.

### Key Trustee Server

KTS is not supported on Cloudera Runtime 7.3.1 and higher.

Component	Memory	CPU	Disk
Key Trustee Server   <b>Note:</b> KTS requires a additional dedicated resources.	8 GB	1 GHz 64-bit quad core	20 GB, using moderate to high-performance drives

### Related Information

[Encrypting Data at Rest](#)

### Ranger KMS

Component	Memory	CPU	Disk
Ranger KMS   <b>Note:</b> Cloudera recommends using machines with CPUs that support the AES-NI instruction set and have a similar performance to the CPUs available to the NameNodes, so as not to introduce a bottleneck to HDFS client operations.	8 GB	1 GHz 64-bit quad core	20 GB, using moderate to high-performance drives

### Kudu

Understand the resource requirements for Kudu before making the resource configuration changes in Cloudera Manager.

Component	Memory	CPU	Disk
Tablet Server	<ul style="list-style-type: none"> <li>Minimum: 4 GB</li> <li>Recommended: 10 GB</li> </ul> <p>Additional hardware may be required, depending on the workloads running in the cluster.</p>	<p>Kudu currently requires a CPU that supports the SSSE3 and SSE4.2 instruction sets.</p> <p>If you are to run Kudu inside a VM, enable SSE4.2 pass-through to pass through SSE4.2 support into the VM.</p>	1 disk for write-ahead log (WAL). Using an SSD drive may improve performance.
Master	<ul style="list-style-type: none"> <li>Minimum: 256 MB</li> <li>Recommended: 1 GB</li> </ul>	<p>Kudu currently requires a CPU that supports the SSSE3 and SSE4.2 instruction sets.</p> <p>If you are to run Kudu inside a VM, enable SSE4.2 pass-through to passthrough SSE4.2 support into the VM.</p>	1 disk

### Related Information

[Apache Kudu configuration](#)

### Oozie

Understand the resource requirements for Oozie before making the resource configuration changes in Cloudera Manager.

Component	Java Heap	CPU	Disk
Oozie	<ul style="list-style-type: none"> <li>Minimum: 1 GB (this is the default set by Cloudera Manager). This is sufficient for less than 10 simultaneous workflows, without forking.</li> <li>If you notice excessive garbage collection, or out-of-memory errors, increase the heap size to 4 GB for medium-size production clusters or to 8 GB for large-size production clusters.</li> <li>Set this value using the Java Heap Size of Oozie Server in Bytes Oozie configuration property.</li> </ul>	No resources required	No resources required

Additional tuning:

For workloads with many coordinators that run with complex workflows (a max concurrency reached! warning appears in the log and the Oozie admin -queuedump command shows a large queue):

- Increase the value of the oozie.service.CallableQueueService.callable.concurrency property to 50.
- Increase the value of the oozie.service.CallableQueueService.threads property to 200.

Do not use a Derby database as a backend database for Oozie.

### Ozone hardware recommendations

This guide helps you choose hardware for Ozone based on your data storage needs. Following these recommendations will ensure that you get optimum performance from your Ozone cluster.

### Table 17: Recommendations

The values are minimums and can go higher depending on your requirements. However, the above recommendations will apply to the vast majority of deployments.

Node Type	Chasis	CPU	Node RAM	RAM for each service	OS Disk	Meta Disk (NVMe)	Data Disk	Network	Disk Controllers	GPU
Master node (OM and SCM and Recon)	1U	2 x 20c	256 GB	64 GB	2 x 480 GB SSD	2 x 4 TB	-	2x 25Gbps	-	-
Datanode (Ozone, no compute)	2U	2 x 12c		31 GB**		2 x 1.5 TB	24 x 16TB	2x 12 Gbps (low)	Optional	
		2 x 24c		512 GB		2 x 3 TB				
Compute node (No Storage)	1U			-		1 x 4 TB	-	1x 25Gbps	-	

\*\*Avoid using heap sizes of 32 GB to 47 GB because the JVM cannot use [Compressed oops](#) for heap sizes > 31 GB. This reduces the effective memory available to the process. If you want to configure heaps > 31 GB, then use a heap size of at least 48 GB or higher.

## Notes

The above configuration will support up to 10B keys because of the 4 TB NVMe on the master nodes.

The absolute minimum recommended configuration is 3 master nodes and 9 datanodes. This will support Erasure Coding with the RS(6,3) configuration with full High Availability. Additional datanodes can be added in increments of 1 to increase storage.

## Network

The network between the datanodes and the compute nodes cannot be oversubscribed by more than 2:1. Networking is sized to support the full (real-world) bandwidth of the drives across the network. More drives require faster networks, both at the server level and the switch level.

## NVMe

NVMe should be configured in RAID1 pairs to provide business continuity for Ozone metadata in case of hardware failure.

The master nodes and datanodes use NVMe to store Ozone metadata. The compute nodes use NVMe for shuffle (Spark, MapReduce, and Tez) and caching (LLAP). The mixed compute datanodes use NVMe for both Ozone metadata and shuffle (Spark, MapReduce, and Tez) plus caching (LLAP).

Cloudera recommends mounting Ozone partitions across the NVMe drive pair as RAID1 (800GB) with the remaining space used for shuffle or cache as independent JBOD partitions. RAID can be configured either in hardware or in software.



### Note:

- Cloudera recommends NVMe for storing OM and SCM metadata.
- Spinning media to store OM or SCM metadata is unsupported.
- For non-production clusters like staging or development clusters, write-optimized SSDs are recommended.

### Example sizing calculator

Suggested Value of Parameter	Logical Capacity 2PB	Logical Capacity 8PB	Logical Capacity 16PB	Additional information
Number of Data Nodes if using Erasure Coding rs(6,3)	9	31	64	These are calculated based on actual file storage required (See Row 1)
Logical data size proposed (TB, EC 6,3)	2304	8192	16384	-
Raw disk capacity (TB)	3456	12288	24576	-
Number of Data Nodes if using triple replication	16	64	128	These are calculated based on actual file storage required (See Row 1)
Logical data size - conservative using 3x (TB)	2048	8192	16384	-
Raw disk capacity (TB)	6144	24576	49152	-

### Related Information

#### Ozone Architecture

### Phoenix

Know the hardware resource requirements for Phoenix Query Server (PQS) in Cloudera Base on premises.

Component	Java Heap	CPU	Disk
Phoenix Query Server	1 GB - 4 GB  Set this value using the Phoenix Query Server Max Heapsize configuration property. Increase this property value if you run any of these queries, aggregates, joins, or subqueries and if the query processing requires more memory.	Minimum 2 dedicated cores.	1 disk for local logs, which can be shared with the operating system and other Hadoop logs.

### Ranger

Memory	CPU	Disk	Additional Dependencies
Ranger Admin: 1 GB minimum, then adjust heap as required (8 GB-16 GB)	1 core minimum	No special requirement.	
Ranger Usersync: 1 GB minimum	1 core minimum	No special requirement.	
Ranger Tagsync: 1 GB minimum	1 core minimum	No special requirement.	

### Solr

Component	Java Heap	CPU	Disk
Solr	<ul style="list-style-type: none"> <li>Small workloads, or evaluations: 16 GB</li> <li>Smaller production environments: 32 GB</li> <li>Larger production environments: 96 GB is sufficient for most clusters.</li> </ul> <p>Set this value using the Java Heap Size of Solr Server in Bytes Solr configuration property.</p>	<ul style="list-style-type: none"> <li>Minimum: 4</li> <li>Recommended: 16 for production workloads</li> </ul>	<p>No requirement if Solr uses HDFS for storage.</p> <p>If Solr uses local file system:</p> <ul style="list-style-type: none"> <li>Faster disks, such as SSD can provide a significant performance improvement.</li> <li>Occasionally a node may need disk space equal to 2-2.5 times the size of shards on the node for storage and overhead (for segment merging and shard recovery).</li> <li>For heavy ingest and query loads Solr typically performs best on dedicated nodes, partially due to available OS cache for Solr files.</li> </ul>

Note the following considerations for determining the optimal amount of heap memory:

- Size of searchable material: The more searchable material you have, the more memory you need. All things being equal, 10 TB of searchable data requires more memory than 1 TB of searchable data.
- Content indexed in the searchable material: Indexing all fields in a collection of logs, email messages, or Wikipedia entries requires more memory than indexing only the Date Created field.
- The level of performance required: If the system must be stable and respond quickly, more memory may help. If slow responses are acceptable, you may be able to use less memory.

## Related Information

[Deployment Planning for Cloudera Search](#)

## Spark

Component	Java Heap	CPU	Disk
Spark History Server	<p>Minimum: 512 MB</p> <p>Set this value using the Java Heap Size of History Server in Bytes Spark configuration property.</p>	<p>1</p>  <p><b>Important:</b> Cloudera recommends that you adjust the number of CPUs and memory for the Spark History Server based on your specific cluster usage patterns.</p>	Minimum 1 disk for log files.

## Livy

Understand the resource requirements for Livy before making the resource configuration changes in Cloudera Manager.

Component	Java Heap	CPU	Disk
Livy	<p>Minimum: 512 MB</p> <p>Set this value using the maximum size for the Java process heap memory Livy configuration property.</p>		Minimum 1 disk

## YARN

Component	Java Heap	CPU	Other Recommendations
Job History Server	<ul style="list-style-type: none"> <li>Minimum: 1 GB</li> <li>Increase memory by 1.6 GB for each 100,000 tasks kept in memory. For example: 5 jobs @ 100, 000 mappers + 20,000 reducers = 600,000 total tasks requiring 9.6 GB of heap.</li> </ul> <p>See the Other Recommendations column for additional tuning suggestions.</p> <p>Set this value using the Java Heap Size of JobHistory Server in Bytes YARN configuration property.</p>	Minimum: 1 core	<ul style="list-style-type: none"> <li>Set the mapreduce.jobhistory.loadedtasks.cache.size property to a total loaded task count. Using the example in the Java Heap column to the left, of 650,000 total tasks, you can set it to 700,000 to allow for some safety margin. This should also prevent the JobHistoryServer from hanging during garbage collection, since the job limit does not have a task limit.</li> </ul>
NodeManager	<p>Minimum: 1 GB.</p> <p>Configure additional heap memory for the following conditions:</p> <ul style="list-style-type: none"> <li>Large number of containers</li> <li>Large shxmlshuffle sizes in Spark or MapReduce</li> </ul> <p>Set this value using the Java Heap Size of NodeManager in Bytes YARN configuration property.</p>	<ul style="list-style-type: none"> <li>Minimum: 8-16 cores</li> <li>Recommended: 32-64 cores</li> </ul>	<p>Disks:</p> <ul style="list-style-type: none"> <li>Minimum: 8 disks</li> <li>Recommended: 12 or more disks</li> </ul> <p>Networking:</p> <ul style="list-style-type: none"> <li>Minimum: Dual 1Gbps or faster</li> <li>Recommended: Single/Dual 10 Gbps or faster</li> </ul>
ResourceManager	<p>Minimum: 6 GB</p> <p>Configure additional heap memory for the following conditions:</p> <ul style="list-style-type: none"> <li>More jobs</li> <li>Larger cluster size</li> <li>Number of retained finished applications (configured with the yarn.resourcemanager.max-completed-applications property).</li> <li>Scheduler configuration</li> </ul> <p>Set this value using the Java Heap Size of ResourceManager in Bytes YARN configuration property.</p>	Minimum: 1 core	
Other Settings	<ul style="list-style-type: none"> <li>Set the ApplicationMaster Memory YARN configuration property to 512 MB</li> <li>Set the Container Memory Minimum YARN configuration property to 1 GB.</li> </ul>	N/A	N/A

## Related Information

[Tuning Apache Hadoop YARN](#)

## ZooKeeper

Understand the resource requirements for ZooKeeper before making the resource configuration changes in Cloudera Manager.

Component	Java Heap	CPU	Disk
ZooKeeper Server	<ul style="list-style-type: none"> <li>• Minimum: 1 GB</li> <li>• Increase heap size when watching 10,000 - 100,000 ephemeral znodes and are using 1,000 or more clients.</li> </ul> <p>Set this value using the Java Heap Size of ZooKeeper Server in Bytes ZooKeeper configuration property.</p>	Minimum: 4 cores	ZooKeeper was not designed to be a low-latency service and does not benefit from the use of SSD drives. The ZooKeeper access patterns – append-only writes and sequential reads – were designed with spinning disks in mind. Therefore Cloudera recommends using HDD drives.

### Related Information

[Add a ZooKeeper service](#)

## Operating System Requirements

This topic describes the operating system requirements for Cloudera Base on premises. Azul OpenJDK, OpenJDK 8, OpenJDK 11, and OpenJDK 17 are TCK certified for Cloudera.

### Cloudera Base on premises Supported Operating Systems

See the [Cloudera Support Matrix](#) for detailed information about supported operating systems.

### Operating System support for the Cloudera Base on premises Trial Installer

SLES 15 SP4 is supported when using the Trial Installer (cloudera-manager-installer.bin) to install Cloudera Manager.



**Important:** Extra step required when using Cloudera Manager Trial installer on SLES 15 SP4.

When using cloudera-manager-installer.bin to install a trial version of Cloudera Manager, the installation will fail.

Before running cloudera-manager-installer.bin, run the following command:

```
SUSEConnect --list-extensions
SUSEConnect -p sle-module-legacy/15.4/x86_64
zypper install libncurses5
```

### Important information about Cloudera Runtime and Cloudera Manager Supported Operating Systems

Cloudera Runtime provides parcels for select versions of RHEL-compatible operating systems.

**Important:**

In order to be covered by Cloudera Support:

- All Cloudera Runtime hosts in a logical cluster must run on the same major OS release.
- Cloudera supports a temporarily mixed OS configuration during an OS upgrade project.
- Cloudera Manager must run on the same OS release as one of the clusters it manages.

Cloudera recommends running the same minor release on all cluster nodes. However, the risk caused by running different minor OS releases is considered lower than the risk of running different major OS releases.

Points to note:

- Cloudera does not support Cloudera Runtime cluster deployments in Docker containers.
- Cloudera Enterprise is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in enforcing mode. Cloudera is not responsible for policy support or policy enforcement. If you experience issues with SELinux, contact your OS provider.

**Important:**

- NavEncrypt is not supported in 7.1.8, 7.1.9, and 7.3.1 when using SLES 15 SP4 or SP5
- Cloudera Manager 7.13.1 supports only SLES 15 but not SLES 12. So it is not possible to have temporarily mixed OS configurations during the upgrade.

**Cloudera Base on premises supported operating systems**

Operating System	Version
IBM PowerPC on RHEL	<p>The following components are not supported:</p> <ul style="list-style-type: none"> <li>• Impala</li> <li>• Kudu</li> <li>• Ozone</li> <li>• Navigator Encrypt</li> </ul> <p> <b>Note:</b> Ranger KMS is the recommended Key Management Server for PowerPC deployments.</p>

**Operating System and IBM PowerPC support matrix**

This matrix explains the operating system supported on IBM PowerPC. There are two core configurations with Cloudera Base on premises and different PowerPC version deployments:

1. IBM PowerPC only and Cloudera Base on premises
2. IBM PowerPC CPU, IBM Spectrum Scale Storage, and Cloudera Base on premises. This is a subset of what is supported generally on IBM PowerPC.

IBM PowerPC Support	Documentation
PowerPC 8 and 9 generally without Spectrum Scale Storage	<a href="https://www.ibm.com/docs/en/linux-on-systems?topic=ipo-supported-linux-distributions-virtualization-options-power8-power9-linux-power-systems">https://www.ibm.com/docs/en/linux-on-systems?topic=ipo-supported-linux-distributions-virtualization-options-power8-power9-linux-power-systems</a>
PowerPC 10 generally without Spectrum Scale Storage	<a href="https://www.ibm.com/docs/en/linux-on-systems?topic=ipo-supported-linux-distributions-virtualization-options-power10-linux-power-servers">https://www.ibm.com/docs/en/linux-on-systems?topic=ipo-supported-linux-distributions-virtualization-options-power10-linux-power-servers</a>
IBM Spectrum Scale Storage with Cloudera Base on premises on x86 and PowerPC combinations	<a href="https://www.ibm.com/docs/en/spectrum-scale-bda?topic=requirements-support-matrix">https://www.ibm.com/docs/en/spectrum-scale-bda?topic=requirements-support-matrix</a>

## Software Dependencies

- Python - Python dependencies for the different Cloudera components is mentioned below:

### Cloudera Manager



**Important:** For information about the supported Python versions for the operating systems, see [Python support matrix](#).

### Hue

The following are the supported Python versions for the operating systems:

Python versions	Supported operating system for CDP Private Cloud Base 7.1.9 SP1 CHF 9 with Cloudera Manager 7.13.1.400	Supported operating system for Cloudera Runtime 7.3.1
Python 3.11	Red Hat 8, Red Hat 9, SLES 15, Ubuntu 22	-
Python 3.10	-	SLES 15, Ubuntu 22
Python 3.9	Red Hat 8, Red Hat 9	Red Hat 8, Red Hat 9
Python 3.8	CentOS 7, Red Hat 8, SLES 12, Ubuntu 20, Ubuntu 22	Ubuntu 20

### Spark



**Important:** Spark 2 is not supported in Cloudera Runtime 7.3.1.



Spark 2.4 supports Python 2.7 and 3.4-3.7.



**Important:** Spark 2 will be deprecated in Cloudera Runtime 7.1.9. Therefore, 7.1.9 is the last runtime release where Spark 2 is supported. For more information, see [Deprecation Notices in Cloudera Runtime](#).

Spark 2.4 supports Python 2.7 and 3.4-3.7.

Spark 3.0 supports Python 2.7 and 3.4 and higher, although support for Python 2 and 3.4 to 3.5 is deprecated.

Spark 3.1 supports Python 3.6 and higher.

If the right level of Python is not picked up by default, set the PYSPARK\_PYTHON and PYSPARK\_DRIVER\_PYTHON environment variables to point to the correct Python executable before running the `pyspark` command.

CDS (Cloudera Distribution of Spark) 3.3 supports Python 3.7 and higher.



**Note:** For a complete list of supported Python versions in different Spark versions, refer to the [Spark Python Supportability Matrix](#).

- Perl - Cloudera Manager requires perl.
- `python-psycopg2` - Cloudera Manager 7 has a dependency on the package `python-psycopg2`. PostgreSQL-backed Hue in Runtime 7 requires a higher version of `psycopg2` than is required by the Cloudera Manager dependency. For more information, see [Installing the psycopg2 Python Package](#).

- **iproute package** - Cloudera Base on premises has a dependency on the iproute package. Any host that runs the Cloudera Manager Agent requires the package. The required version varies depending on the operating system:

**Table 18: iproute package**

Operating System	iproute version
RHEL	iproute
Ubuntu	iproute2
SLES	iproute2

- **rpcbind package** - Cloudera Base on premises has a dependency on the `rpcinfo` command which is usually found in the rpcbind package. Any host that runs the Cloudera Manager Agent requires this package. The required version varies depending on the operating system.

## Filesystem Requirements

The Hadoop Distributed File System (HDFS) is designed to run on top of an underlying filesystem in an operating system.

### Supported Filesystems

Cloudera recommends that you use either of the following filesystems tested on the supported operating systems:

- ext3: This is the most tested underlying filesystem for HDFS.
- ext4: This scalable extension of ext3 is supported in more recent Linux releases.



**Important:** Cloudera does not support in-place upgrades from ext3 to ext4. Cloudera recommends that you format disks as ext4 before using them as data directories.

- XFS: This is the default filesystem in RHEL.
- S3: Amazon Simple Storage Service

Kudu Filesystem Requirements - Kudu is supported on ext4 and XFS. Kudu requires a kernel version and filesystem that supports hole punching. Hole punching is the use of the `fallocate(2)` system call with the `FALLOC_FL_PUNCH_HOLE` option set.

### File Access Time

Linux filesystems keep metadata that record when each file was accessed. This means that even reads result in a write to the disk. To speed up file reads, Cloudera recommends that you disable this option, called `atime`, using the `noatime` mount option in `/etc/fstab`:

```
/dev/sdb1 /data1 ext4 defaults,noatime 0
```

Apply the change without rebooting:

```
mount -o remount /data1
```

### Filesystem Mount Options

The filesystem mount options have a `sync` option that allows you to write synchronously.

Using the `sync` filesystem mount option reduces performance for services that write data to disks, such as HDFS, YARN, Kafka and Kudu. In Cloudera, most writes are already replicated. Therefore, synchronous writes to disk are unnecessary, expensive, and do not measurably improve stability.

NFS and NAS options are not supported for use as DataNode Data Directory mounts, even when using Hierarchical Storage features.

Cloudera supports mounting /tmp with the noexec option. Mounting /tmp as a filesystem with the noexec option is sometimes done as an enhanced security measure to prevent the execution of files stored there.

## Filesystem Requirements

You can control resource allocation for Cloudera Manager and Cloudera Runtime services (nproc, nofile, etc) from /etc/security/limits.conf, and through init scripts on traditional SysV Init systems. However, on systems using systemd the limits either needs to be set in the service's unit file, or in /etc/systemd/system.conf, or in files present under /etc/systemd/system.conf.d/\*. This is due to a known limitation with systemd as it does not use PAM login sessions (pam\_limits.so) for daemon services, thereby ignoring the limits defined in /etc/security/limits.conf. Both Cloudera Manager Agent and Supervisord (responsible for starting Cloudera Runtime services) are daemonised during system initialisation.

You can perform either of the following steps to modify the resource limit:

1. For system-wide change, uncomment the process properties from /etc/systemd/system.conf, or create an override .conf under /etc/systemd/system.conf.d/. This requires a \*nix system reboot for the changes to take effect. For more information, see [Limits.conf](#).
2. To apply custom limits on Cloudera Runtime services, add the required process properties to the [Service] section in /usr/lib/systemd/system/cloudera-scm-supervisord.service.

For instance, to customise the number of child processes a process can fork. You can set the property as follows:

```
LimitNPROC=<value>
```

Then reload the configuration by running the following command for the limits to be applied in the subsequent service restarts:

```
# systemctl daemon-reload
```

Here are the list of available [process properties](#).

## nscd for Kudu

Although not a strict requirement, it's highly recommended that you use nscd to cache both DNS name resolution and static name resolution for Kudu.

## Configuring system level operating system

Cloudera recommends you to set up the following configurations:

- Disabling Transparent Hugepages (THP)
- vm.swappiness Linux Kernel Parameter

For setting these configurations, see [Disabling Transparent Hugepages \(THP\)](#) and [Setting the vm.swappiness Linux Kernel Parameter](#).

## Database Requirements

This topic describes the database requirements for Cloudera Base on premises.

See [Cloudera Support Matrix](#) for detailed information about supported databases based on the Cloudera and Cloudera Manager versions.



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

Cloudera Manager and Cloudera Runtime come packaged with an embedded PostgreSQL database for use in non-production environments. The embedded PostgreSQL database is not supported in production environments. For production environments, you must configure your cluster to use dedicated external databases. You must ensure latency between Cloudera Manager server and the database is < 10 ms. You can verify the latency with a simple SQL command from your Cloudera Manager server host to the database. Start your database's command line client tool and connect to the Cloudera Manager database. Run the SQL command `SELECT 1;`



**Important:** Migrating from one external database server to a different type of database server after installing Cloudera Manager and Cloudera Runtime services is a complex process that requires modification of the schema and matching the data in the database tables to the new schema. Cloudera expects you to perform the migration using any off-the-shelf tool. If you require any assistance, Cloudera recommends that you must engage with Cloudera Professional Services team.

After installing a database, upgrade to the latest patch and apply appropriate updates. Available updates may be specific to the operating system on which it is installed.

Notes:

- Cloudera recommends installing the databases on different hosts than the services, located in the same data center. 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.
- Hue Query Processor in CDP Private Cloud Base 7.1.8 requires a non-SSL enabled PostgreSQL database.
- Cloudera does not support Percona for MySQL as a backend database for Hive Metastore (HMS).
- Use the appropriate UTF8 encoding for Metastore, Oozie, Hive, and Hue.

Oozie also supports UTF8MB4 character encoding out of box without any configuration change when the Oozie custom database is created with the encoding of UTF8MB4.

MySQL and MariaDB must use the MySQL utf8 encoding, not utf8mb4.

- Ranger only supports the InnoDB engine for MySQL and MariaDB databases.
- YARN Queue Manager requires a Postgres database server with a dedicated database.
- For MySQL 5.7, you must install the MySQL-shared-compat or MySQL-shared package. This is required for the Cloudera Manager Agent installation.
- MySQL GTID-based replication is not supported.
- Both the Community and Enterprise versions of MySQL are supported, as well as MySQL configured by the AWS RDS service.
- Before upgrading from CDH 5 to CDH 6, check the value of the COMPATIBLE initialization parameter in the Oracle Database using the following SQL query:

```
SELECT name, value FROM v$parameter WHERE name = 'compatible'
```

The default value is 12.2.0. If the parameter has a different value, you can set it to the default as shown in the [Oracle Database Upgrade Guide](#).



**Note:** Before resetting the COMPATIBLE initialization parameter to its default value, make sure you consider the effects of this change can have on your system.

## RDBMS High Availability Support

Various Cloudera components rely on backing RDBMS services as critical infrastructure. You may require Cloudera components to support deployment in environments where RDBMS services are made highly-available. High availability (HA) solutions for RDBMS are implementation-specific, and can create constraints or behavioral changes in Cloudera components.

This section clarifies the support state and identifies known issues and limitations for HA deployments.

### Upgrading Cloudera Manager and the Cloudera Manager database

When upgrading Cloudera Manager, there may be a minimum version requirement for the database server.

Ensure that the Cloudera Manager database server is upgraded to at least this minimum requirement prior to starting the new version of Cloudera Manager for the first time.

1. Stop Cloudera Management Service.
2. Upgrade the Cloudera Manager RPMs.
3. Upgrade the Cloudera Manager database server version.
4. Start the Cloudera Manager.

### **Upgrading Cloudera Base on premises and the Cloudera Base on premises Services databases**

When upgrading Cloudera Base on premises to a new version, the new version of Cloudera Base on premises may have a minimum version requirement for the database server which holds services metadata.

If the database server needs to be upgraded, follow this process:

1. Stop Cloudera Base on premises services which depend on the database server, or alternatively, stop the entire cluster.
2. Upgrade the database server version.
3. Upgrade Cloudera Base on premises using Cloudera Manager.
4. Start the stopped services, or the entire cluster.

When a database server upgrade is required, it is not possible to perform a rolling upgrade of the cluster.

### **Difference between High Availability and Load Balancing**

Understanding the difference between HA and load balancing is important for Cloudera components, which are designed to assume services are provided by a single RDBMS instance. Load balancing distributes operations across multiple RDBMS services in parallel, while HA focuses on service continuity. Load balanced deployments are often used as part of HA strategies to overcome demands of monitoring and failover management in an HA environment. While less easier to implement, load-balanced deployments require applications tailored to the behavior and limitations of the particular technology.

Support Statement: Cloudera components are not designed for and do not support load balanced deployments of any kind. Any HA strategy involving multiple active RDBMS services must ensure all connections are routed to a single RDBMS service at any given time, regardless of vendor or HA implementation/technology.

### **General High Availability Support**

Cloudera supports various RDBMS options, each of which have multiple possible strategies to implement HA. Cloudera cannot reasonably test and certify on each strategy for each RDBMS. Cloudera expects HA solutions for RDBMS to be transparent to Cloudera software, and therefore are not supported and debugged by Cloudera. It is the responsibility of the customer to provision, configure, and manage the RDBMS HA deployment, so that Cloudera software behaves as it would when interfacing with a single, non-HA service. Cloudera will support and help customers troubleshoot issues when a cluster has HA enabled. While diagnosing database-related problems in Cloudera components, customers may be required to temporarily disable or bypass HA mechanisms for troubleshooting purposes. If an HA-related issue is found, it is the responsibility of the customer to engage with the database vendor so that a solution to that issue can be found.

Support Statement: Cloudera Support may require customers to temporarily bypass HA layers and connect directly to supported RDBMS back-ends to troubleshoot issues. Issues observed only when connected through HA layers are the responsibility of the customer DBA staff to resolve.

### **RDBMS Storage Sizing**

The amount of RDBMS storage space used by Cloudera Base on premises varies depending on the services that are installed and the operations performed. Approximately, the amount of RDBMS storage needed is between 10 MB and 100 MB per host in the Cloudera cluster.

You can better estimate the RDBMS storage space by deploying a test cluster with the approximate proportion of service roles that the full cluster can bear. Later, execute a sample set of operations, (including Data Recovery backup) for about 24 hours and observe the storage usage on the RDBMS. Next, extrapolate the usage to the full cluster size.

## Sharing an RDBMS with other applications

The ability to share an RDBMS storage between Cloudera Base on premises and other applications depends on many factors. Cloudera recommends that you do not share the RDBMS used by Cloudera Base on premises with any other application.

For non-production clusters where cluster size is small, not expected to grow, and occasional glitches are tolerable, it is acceptable to share a database with other applications.

### MySQL

For a production cluster, Cloudera Runtime services must not share a database server with other applications. For small clusters, this database can be shared by the Cloudera Runtime services. For large clusters (hosts > 500), each Cloudera Runtime service must have its own database server.

### PostgreSQL

If you have a dedicated database team managing high-performance hardware, with the Cloudera Base on premises databases stored on their own spindles (or raid array), then it can be possible to have the DB server shared with other applications. When the cluster size is very large (hosts > 1000), there might be performance issues between shared applications. Cloudera recommends that you do not share the Cloudera Base on premises database server with other application usages.

If you do not have a dedicated database team that can analyze and tune RDBMS performance, it is recommended to follow the advice for MySQL as detailed above.

### Oracle

For single-server Oracle installations, see the above description related to PostgreSQL.

If you are using a clustered system like Oracle RAC, with multiple servers, it is possible to use a shared DB service, since it is no longer a single server. The end user's DB team must monitor DB latency, scale the hardware, or tune DB parameters to ensure optimal performance.

### Latency target

For end users attempting to tune a shared RDBMS, ensure that elapsed times must not exceed 40 milliseconds for the 99th percentile of SELECT statements on indexed single-row queries.

## Java Requirements

Cloudera Base on premises comprises of Cloudera Manager and Cloudera Runtime services. Understand the specific JDK requirements for your deployments.

### Supported JDKs

Please see the [Cloudera Support Matrix](#) for detailed information about supported JDKs.

For Cloudera Runtime services, Open JDK-17 is provided as an additional platform support for Cloudera Runtime 7.1.9.

### Supported JDK versions

**Warning:**

- Upgrading to Oracle JDK 1.8.351 or higher, OpenJDK 11.0.17 or higher, or OpenJDK 1.8.392, and Spark3 in CDE uses OpenJDK 11.0.17 which causes a Kerberos issue when deprecated 3DES and RC4 permitted encryption types are used.

Workaround: Remove the deprecated 3DES and RC4 encryption types in the krb5.conf and kdc.conf files.

- JDK 8u271, JDK 8u281, and JDK 8u291 may cause socket leak issues due to JDK-8245417 and JDK-8256818. Pay attention to the build version of your JDK because some later builds are fixed as described in [JDK-8256818](#).

Workaround: Consider using a more recent version of the JDK like 8u282, or builds of the JDK where the issue is fixed.

- [Oozie Workflow Graph Display](#) in Hue does not work properly with JDK versions lower than 8u40.
- The default value of the YARN config **Add add-opens flags to MR containers** is **false** regardless of the JDK, Cloudera Manager or Cloudera version. To run distcp or any MapReduce application, this config must be turned on whenever JDK17 is used.
- If you are using JDK 17 on your cluster, you must run the JVM option `--add-opens=java.base/java.lang=ALL-UNNAMED --add-opens=java.management/com.sun.jmx.mbeanserver=ALL-UNNAMED --add-exports=java.management/com.sun.jmx.mbeanserver=ALL-UNNAMED --add-exports=java.base/sun.net.dns=ALL-UNNAMED --add-exports=java.base/sun.net.util=ALL-UNNAMED` to ensure the jobs run successfully.

**Table 19: Azul Open JDK versions that are tested and recommended**

Azul Open JDK Version	Notes
17.0.7	
11.50.19	
8.56.0.21	Minimum required version

**Table 20: Oracle JDK versions that are tested and recommended**

Oracle JDK Version	Notes
17.0.6	
11.0.10+8	
1.8u181	Minimum required version

**Table 21: OpenJDK versions that are tested and recommended**

OpenJDK Version	Notes
17.0.7	17.0.2 is the minimum required OpenJDK 17 version for FIPS
11.0.4+11	11.0.3 is the minimum required OpenJDK 11 version for FIPS
1.8u231	For FIPS minimum required / latest version tested
1.8u232	Minimum required / Latest version tested



**Note:** Note the following about OpenJDK support:

- Updates above the minimum that are not listed are supported but not tested.
- Cloudera tests only the OpenJDK builds that are provided by each operating system, and only the versions listed in the support matrix.



**Note:** Cloudera Manager supports TLS 1.2 for Java 8 and Java 11. For Java 17 and higher versions, Cloudera Manager supports TLS 1.2 and TLS 1.3. For TLS 1.0 and TLS 1.1, Cloudera Manager supports Java 8, though Cloudera recommends not to use TLS 1.0 and TLS 1.1.



#### Warning:

- Spark3 in Cloudera Data Engineering uses OpenJDK 11.0.17 which causes a Kerberos issue when deprecated 3DES and RC4 permitted encryption types are used.

Workaround: Remove the deprecated 3DES and RC4 encryption types in the krb5.conf and kdc.conf files.

- JDK 8u271, JDK 8u281, and JDK 8u291 may cause socket leak issues due to JDK-8245417 and JDK-8256818. Pay attention to the build version of your JDK because some later builds are fixed as described in [JDK-8256818](#).

Workaround: Consider using a more recent version of the JDK like 8u282, or builds of the JDK where the issue is fixed.

- Upgrading to Oracle JDK 1.8.351 or higher, OpenJDK 11.0.17 or higher, or OpenJDK 1.8.392 causes a Kerberos issue when deprecated 3DES and RC4 permitted encryption types are used. You can workaround this issue by removing the deprecated 3DES and RC4 encryption types in the krb5.conf and kdc.conf files.
- Spark3 in Cloudera Data Engineering uses OpenJDK 11.0.17 which causes a Kerberos issue when deprecated 3DES and RC4 permitted encryption types are used.

Workaround: Remove the deprecated 3DES and RC4 encryption types in the krb5.conf and kdc.conf files.

- JDK 8u40, 8u45, and 8u60 are not supported due to JDK issues impacting CDH functionality:
  - JDK 8u40 and 8u45 are affected by [JDK-8077155](#), which affects HTTP authentication for certain web UIs.
  - JDK 8u60 is incompatible with the AWS SDK, and causes problem with DistCP. For more information, see the [KB article](#).
- [Oozie Workflow Graph Display](#) in Hue does not work properly with JDK versions lower than 8u40.
- The default value of the YARN config **Add add-opens flags to MR containers** is **false** regardless of the JDK, Cloudera Manager or Cloudera version. To run distcp or any MapReduce application, this config must be turned on whenever JDK17 is used.
- If you are using JDK 17 on your cluster, you must run the JVM option `--add-opens=java.base/java.lang=ALL-UNNAMED --add-opens=java.management/com.sun.jmx.mbeanserver=ALL-UNNAMED --add-exports=java.management/com.sun.jmx.mbeanserver=ALL-UNNAMED --add-exports=java.base/sun.net.dns=ALL-UNNAMED --add-exports=java.base/sun.net.util=ALL-UNNAMED` to ensure the jobs run successfully.



#### Important:

For JDK 8u241 and higher versions running on Kerberized clusters, you must disable referrals by setting `sun.security.krb5.disableReferrals=true`.

For example, with OpenJDK 1.8.0u242:

1. Open `/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.242.b08-0.el7_7.x86_64/jre/lib/security/java.security` with a text editor.
2. Add `sun.security.krb5.disableReferrals=true` (it can be at the bottom of the file).
3. Add this property on each node that has the impacted JDK version.
4. Restart the applications using the JDK so the change takes effect.

For more information, see the [KB article](#).

## Support Notes



**Note:** Cloudera recommends installing Oracle JDK at /usr/java/<JDK-VERSION> and OpenJDK at /usr/lib/jvm, which allows Cloudera Manager to auto-detect and use the correct JDK version. If you install the JDK anywhere else, there are additional steps required to configure Cloudera Manager with your chosen location. See [Configuring a custom Java Home Location](#).



**Note:** A Java optimization called compressed oops (ordinary object pointers) enables a 64-bit JVM to address heap sizes up to about 32 GB using 4-byte pointers. For larger heap sizes, 8-byte pointers are required. This means that a heap size slightly less than 32 GB can hold more objects than a heap size slightly more than 32 GB.

If you do not need more than 32 GB heap, set your heap size to 31GB or less to avoid this issue. If you need 32 GB or more, set your heap size to 48 GB or higher to account for the larger pointers. In general, for heap sizes above 32 GB, multiply the amount of heap you need by 1.5.

Only 64 bit JDKs are supported.

Unless specifically excluded, Cloudera supports later updates to a major JDK release from the release that support was introduced. Cloudera excludes or removes support for select Java updates when security is jeopardized.

Running Cloudera Runtime nodes within the same cluster on different JDK releases is not supported. All cluster hosts must use the same JDK update level.

## Networking and Security Requirements

This topic describes the networking and security requirements for Cloudera Base on premises.

### Cloudera Runtime and Cloudera Manager Supported Transport Layer Security Versions



**Note:**

- Cloudera Manager supports TLS 1.2 for Java 8 and Java 11. For Java 17 and higher versions, Cloudera Manager supports TLS 1.2 and TLS 1.3. For TLS 1.0 and TLS 1.1, Cloudera Manager supports Java 8, though Cloudera recommends not to use TLS 1.0 and TLS 1.1.
- Domain name requirement: Environments, where domain names have non-LDH characters such as letters, digits, or hyphens, must be avoided for Hadoop deployments as they are unsupported.

The following components are supported by the indicated versions of Transport Layer Security (TLS):

**Table 22: Components Supported by TLS**

Component	Role	Name	Port	Version
Cloudera Manager	Cloudera Management Service		7182	TLS 1.2
Cloudera Manager	Cloudera Management Service		7183	TLS 1.2
Flume			9099	TLS 1.2
Flume		Avro Source/Sink		TLS 1.2
Flume		Flume HTTP Source/Sink		TLS 1.2
HBase	Master	HBase Master Web UI Port	60010	TLS 1.2
HDFS	NameNode	Secure NameNode Web UI Port	50470	TLS 1.2
HDFS	Secondary NameNode	Secure Secondary NameNode Web UI Port	50495	TLS 1.2
HDFS	HttpFS	REST Port	14000	TLS 1.1, TLS 1.2
Hive	HiveServer2	HiveServer2 Port	10000	TLS 1.2
Hue	Hue Server	Hue HTTP Port	8888	TLS 1.2

Component	Role	Name	Port	Version
Impala	Impala Daemon	Impala Daemon Beeswax Port	21000	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala Daemon	Impala Daemon HiveServer2 Port	21050	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala Daemon	Impala Daemon Backend Port	22000	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala StateStore	StateStore Service Port	24000	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala Daemon	Impala Daemon HTTP Server Port	25000	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala StateStore	StateStore HTTP Server Port	25010	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala Catalog Server	Catalog Server HTTP Server Port	25020	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Impala	Impala Catalog Server	Catalog Server Service Port	26000	TLS 1.0, TLS 1.1, TLS 1.2 We recommend that clients use the highest supported version, TLS 1.2.
Oozie	Oozie Server	Oozie HTTPS Port	11443	TLS 1.1, TLS 1.2
Ranger	Ranger Admin	Admin HTTPS Port	6182	TLS 1.2
Ranger KMS	Ranger KMS Server	Ranger KMS HTTPS Port	9494	TLS 1.2
Solr	Solr Server	Solr HTTP Port	8983	TLS 1.1, TLS 1.2
Solr	Solr Server	Solr HTTPS Port	8985	TLS 1.1, TLS 1.2
Spark	History Server		18080	TLS 1.2
YARN	ResourceManager	ResourceManager Web Application HTTP Port	8090	TLS 1.2
YARN	JobHistory Server	MRv1 JobHistory Web Application HTTP Port	19890	TLS 1.2

### Cloudera Runtime and Cloudera Manager Networking and Security Requirements

The hosts in a Cloudera Manager deployment must satisfy the following networking and security requirements:

- Networking Protocols Support

Cloudera requires IPv4. IPv6 is not supported and must be disabled.



**Important:** Refer to your OS documentation or contact your OS vendor for instructions on disabling IPv6.

See also *Configure Network Names*.

- Multihoming Support

Multihoming Cloudera Runtime or Cloudera Manager is not supported outside specifically certified Cloudera partner appliances such as Oracle Big Data Appliance, Teradata Appliance for Hadoop, and Cray® Urika-XA system. Cloudera finds that current Hadoop architectures combined with modern network infrastructures and security practices remove the need for multihoming. Multihoming, however, is beneficial internally in appliance form factors to take advantage of high-bandwidth InfiniBand interconnects.

Although some subareas of the product may work with unsupported custom multihoming configurations, there are known issues with multihoming. In addition, unknown issues may arise because multihoming is not covered by our test matrix outside the Cloudera-certified partner appliances.

- Entropy

Data at rest encryption requires sufficient entropy to ensure randomness.

See entropy requirements in *Data at Rest Encryption Requirements*.

- Cluster hosts must have a working network name resolution system and correctly formatted /etc/hosts file. All cluster hosts must have properly configured forward and reverse host resolution through DNS. The /etc/hosts files must:

- Contain consistent information about hostnames and IP addresses across all hosts
- Not contain uppercase hostnames
- Not contain duplicate IP addresses

Cluster hosts must not use aliases, either in /etc/hosts or in configuring DNS. A properly formatted /etc/hosts file should be similar to the following example:

```
127.0.0.1 localhost.localdomain localhost
192.168.1.1 cluster-01.example.com cluster-01
192.168.1.2 cluster-02.example.com cluster-02
192.168.1.3 cluster-03.example.com cluster-03
```

- In most cases, the Cloudera Management Service must have SSH access to the cluster hosts when you run the installation or upgrade wizard. You must log in using a root account or an account that has password-less sudo permission. For authentication during the installation and upgrade procedures, you must either enter the password or upload a public and private key pair for the root or sudo user account. If you want to use a public and private key pair, the public key must be installed on the cluster hosts before you use Cloudera Manager.

Cloudera Manager uses SSH only during the initial install or upgrade. Once the cluster is set up, you can disable root SSH access or change the root password. Cloudera Manager does not save SSH credentials, and all credential information is discarded when the installation is complete.

- The Cloudera Manager Agent runs as root so that it can make sure that the required directories are created and that processes and files are owned by the appropriate user (for example, the hdfs and mapred users).
- Security-Enhanced Linux (SELinux) must not block Cloudera Manager or Cloudera Runtime operations.



**Note:** Cloudera Enterprise 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.

- Firewalls (such as iptables and firewalld) must be disabled or configured to allow access to ports used by Cloudera Manager, Cloudera Runtime, and related services.
- For RHEL and CentOS, the /etc/sysconfig/network file on each host must contain the correct hostname.
- Cloudera Manager and Cloudera Runtime use several user accounts and groups to complete their tasks. The set of user accounts and groups varies according to the components you choose to install. Do not delete these accounts or groups and do not modify their permissions and rights. Ensure that no existing systems prevent these accounts and groups from functioning. For example, if you have scripts that delete user accounts not in an allowlist, add these accounts to the list of permitted accounts. Cloudera Manager, Cloudera Runtime, and managed services create and use the following accounts and groups:

**Table 23: Users and Groups**

Component (Version)	Unix User ID	Groups	Functionality
Apache Atlas	atlas	atlas, hadoop	Apache Atlas by default has atlas as user and group. It is configurable
Apache Flink	flink	flink	The Flink Dashboard runs as this user.
Apache HBase	hbase	hbase	The Master and the RegionServer processes run as this user.
Apache HBase Indexer	hbase	hbase	The indexer servers are run as this user.
Apache HDFS	hdfs	hdfs, hadoop	The NameNode and DataNodes run as this user, and the HDFS root directory as well as the directories used for edit logs should be owned by it.
Apache Hive Hive on Tez	hive	hive	The HiveServer2 process and the Hive Metastore processes run as this user. A user must be defined for Hive access to its Metastore DB (for example, MySQL or Postgres) but it can be any identifier and does not correspond to a Unix uid. This is javax.jdo.option.ConnectionUserName in hive-site.xml.
Apache Impala	impala	impala, hive	Impala services run as this user.
Apache Kafka	kafka	kafka	Kafka brokers, mirrorMaker, and Connect workers run as this user.
Apache Knox	knox	knox	Apache Knox Gateway Server runs as this user
Apache Kudu	kudu	kudu	Kudu services run as this user.
Apache Livy	livy	livy	The Livy Server process runs as this user
Apache NiFi	nifi	nifi	Runs as the nifi user
Apache NiFi Registry	nifiregistry	nifiregistry	Runs as the nifiregistry user
Apache Oozie	oozie	oozie	The Oozie service runs as this user.
Apache Ozone	hdfs	hdfs, hadoop	Ozone Manager, Storage Container Manager (SCM), Recon and Ozone Datanodes run as this user.
Apache Parquet	~	~	No special users.

Component (Version)	Unix User ID	Groups	Functionality
Apache Phoenix	phoenix	phoenix	The Phoenix Query Server runs as this user
Apache Ranger	ranger	ranger, hadoop	Ranger Admin, Usersync and Tagsync services by default have ranger as user and ranger, hadoop as groups. It is configurable.
Apache Ranger KMS	kms	kms	Ranger KMS runs with kms user and group. It is configurable.
Apache Ranger Raz	rangerraz	ranger	Ranger Raz runs with rangerraz user and is part of the ranger group.
Apache Ranger RMS	rangerrms	ranger	Ranger RMS runs with rangerrms user and is part of the ranger group.
Apache Solr	solr	solr	The Solr processes run as this user.
Apache Spark	spark	spark	The Spark History Server process runs as this user.
Apache Sqoop	sqoop	sqoop	This user is only for the Sqoop1 Metastore, a configuration option that is not recommended.
Apache YARN	yarn	yarn, hadoop	Without Kerberos, all YARN services and applications run as this user. The LinuxContainerExecutor binary is owned by this user for Kerberos.
Apache Zeppelin	zeppelin	zeppelin	The Zeppelin Server process runs as this user
Apache ZooKeeper	zookeeper	zookeeper	The ZooKeeper processes run as this user. It is not configurable.
Cloudera Manager (all versions)	cloudera-scm	cloudera-scm	Clusters managed by Cloudera Manager run Cloudera Manager Server, monitoring roles, and other Cloudera Server processes as cloudera-scm. Requires keytab file named cmf.keytab because name is hard-coded in Cloudera Manager.
Cruise Control	cruisecontrol	hadoop	The Cruise Control process runs as this user.
HttpFS	httpfs	httpfs	The HttpFS service runs as this user. See “HttpFS authentication” for instructions on how to generate the merged httpfs-http.keytab file.
Hue	hue	hue	Hue services run as this user.
Hue Load Balancer	apache	apache	The Hue Load balancer has a dependency on the apache2 package that uses the apache user name. Cloudera Manager does not run processes using this user ID.
Key Trustee Server	keytrustee	keytrustee	The Key Trustee Server service runs as this user.

Component (Version)	Unix User ID	Groups	Functionality
Schema Registry	schemaregistry	hadoop	The Schema Registry process runs as this user.
Streams Messaging Manager	streamsmgr	streamsmgr	The Streams Messaging Manager processes runs as this user.
Streams Replication Manager	streamsrepmgr	streamsrepmgr	The Streams Replication Manager processes runs as this user.

## Data at Rest Encryption Requirements

This topic describes the data at rest encryption requirements for Cloudera Base on premises.

Encryption comprises several components, each with its own requirements.

Data at rest encryption protection can be applied at a number of levels within Hadoop:

- OS filesystem-level
- Network-level
- HDFS-level (protects both data at rest and in transit)

This section contains the various hardware and software requirements for all encryption products used for Data at Rest Encryption.

For more information on supported operating systems, see [Cloudera Support Matrix](#).

For more information on the components, concepts, and architecture for encrypting data at rest, see [Encrypting Data at Rest](#).

## Third-party filesystems

This topic describes the third-party filesystems supported by Cloudera Base on premises.

Cloudera Base on premises supports the following third-party filesystems:

### Third-party filesystem support: Dell EMC PowerScale

Requirements for Dell EMC Power scale (OneFS).

Cloudera Manager version	Cloudera Runtime version	Notes
7.13.1 CHF4	7.1.9 SP1 CHF10	Powerscale Version 9.10
7.13.1	7.3.1	PowerScale version 9.7 ECS version 3.8 (ECS as SFS with HDFS as primary)
7.11.3 CHF7	7.1.9 SP1	PowerScale version 9.7 ECS version 3.8 (ECS as SFS with HDFS as primary)
7.11.3 CHF5	7.1.9 CHF6	PowerScale version 9.5
7.11.3 CHF4	7.1.9 CHF4	PowerScale version 9.5
7.11.3 CHF2	7.1.9 CHF2	PowerScale version 9.5
7.11.3 CHF1	7.1.9 CHF1	PowerScale version 9.5 ECS version 3.8 (ECS as SFS with HDFS as primary)
7.11.3 CHF7	7.1.7 SP3 CHF2	PowerScale version 9.7

Cloudera Manager version	Cloudera Runtime version	Notes
7.11.3 CHF6	7.1.7 SP3 CHF1	PowerScale version 9.7
7.11.3 CHF5	7.1.7 SP3	PowerScale version 9.5 ECS version 3.8 (ECS as SFS with HDFS as primary)
7.6.7 CHF11	7.1.7 SP2 CHF17	PowerScale version 9.5
7.6.7	7.1.7 SP2	PowerScale version 9.3 ECS version 3.7 (ECS as SFS with HDFS as primary)
7.7.1 CHF18	7.1.8 CHF20	PowerScale version 9.5
7.7.1 CHF17	7.1.8 CHF18	PowerScale version 9.5
7.7.1 CHF16	7.1.8 CHF17	PowerScale version 9.5
7.7.1	7.1.8	PowerScale version 9.3 ECS version 3.7 (ECS as SFS with HDFS as primary)
7.6.1	7.1.7 SP1	PowerScale versions 9.2, 9.3 ECS version: 3.6.2.0 (ECS as SFS with HDFS as primary)
7.4.4	7.1.7	OneFS 8.2.2
7.3.1	7.1.6	OneFS 8.2.2 See <a href="#">Hadoop Distributions and Products Supported by OneFS</a>

**Table 24: Dell EMC PowerScale for Cloudera on premises data services**

Cloudera Manager version	Cloudera Runtime version	Cloudera On Premises Data Services version	Notes
7.11.3.28 CHF11	7.1.9.1023 SP1 CHF4	1.5.4 SP1	PowerScale version 9.9
7.11.3.11 CHF 7	7.1.9 SP1	1.5.4	PowerScale version 9.7

Support notes for Dell EMC PowerScale:

- Upgrade from CDP Private Cloud Base 7.1.9 SP1 (CM-7.11.3 CHF7) to CDP Private Cloud Base 7.1.9 SP1 CHF10 (CM-7.13.1 CHF4) is supported.
- Upgrade from CDP Private Cloud Base 7.1.9 SP1 (CM-7.11.3 CHF7) to Cloudera Base on premises 7.3.1 (CM-7.13.1) is supported.
- Upgrade from CDP Private Cloud Base 7.1.8 (CM-7.7.3 CHF2) to CDP Private Cloud Base 7.1.9 CHF1 is supported.
- Upgrade from CDP Private Cloud Base 7.1.7 SP2 (CM-7.6.7) to CDP Private Cloud Base 7.1.9 CHF1 is supported.
- Upgrade from CDP Private Cloud Base 7.1.7 to CDP Private Cloud Base 7.1.8 is supported.
- Upgrades from CDP Private Cloud Base 7.1.6 and 7.1.7 to CDP Private Cloud Base 7.1.7 SP1 are supported.
- Upgrades from CDH 5 and HDP 2 to CDP Private Cloud Base 7.1.7 are supported. For upgrade procedures, see the following documents from Dell EMC:
  - [PowerScale: CDH5 to CDP Private Cloud Base upgrade process](#)
  - [PowerScale: HDP2 to CDP Private Cloud Base upgrade process](#)
- Upgrades from CDH 6 and HDP 3 to CDP Private Cloud Base 7.1.7 are supported. For upgrade procedures, see the following documents from Dell EMC :
  - [PowerScale: CDH6 to CDP Private Cloud Base upgrade process](#)
  - [PowerScale: HDP3 to CDP Private Cloud Base upgrade process](#)

- To configure TLS for Cloudera Runtime components, use Cloudera Manager to configure Auto-TLS. See [Configuring TLS Encryption for Cloudera Manager Using Auto-TLS](#)
- Cloudera Data Catalog for Cloudera Base on premises supports the Dell EMC PowerScale filesystem only when set up with HTTPS.

## Third-party filesystem support: IBM Spectrum Scale

Requirements for IBM Spectrum Scale

Cloudera Manager version	Cloudera Runtime version	Notes
7.11.3 CHF2	7.1.9 CHF2	Supported with IBM Spectrum Scale version 5.1.8
7.11.3 CHF1	7.1.9 CHF1	Supported with IBM Spectrum Scale version 5.1.8
7.7.1 CHF18	7.1.8 CHF 20	Supported with IBM Spectrum Scale version 5.1.8
7.7.1 CHF17	7.1.8 CHF18	Supported with IBM Spectrum Scale version 5.1.8
7.7.1 CHF16	7.1.8 CHF17	Supported with IBM Spectrum Scale version 5.1.8
7.7.1	7.1.8	Supported with IBM Spectrum Scale version 5.1.4
7.11.3 CHF4	7.1.7 SP3	Supported with IBM Spectrum Scale version 5.1.8
7.6.7 CHF11	7.1.7 SP2 CHF17	Supported with IBM Spectrum Scale version 5.1.8
7.6.1	7.1.7 SP1	Supported with RHEL 7.9, 8.2, and 8.4 In-place upgrades to CDP Private Cloud Base 7.1.7 SP1 are only supported from CDP Private Cloud Base 7.1.7 IBM Spectrum Scale versions <sup>1</sup> : 5.1.2.2 and higher
7.4.4	7.1.7	Supported with RHEL 7.9 and 8.2 IBM Spectrum Scale versions <sup>1</sup> : 5.1.2.2 and higher
7.3.1	7.1.6	Supported with RHEL 7.7 and 7.9 IBM Spectrum Scale versions <sup>1</sup> : 5.1.1.0 - 5.1.1.1 Limitation: In-place upgrades from Cloudera Manager 7.2.3 and Cloudera Runtime 7.1.4 to Cloudera Manager 7.3.1 and Cloudera Runtime 7.1.6 are not supported with IBM Spectrum Scale.
7.2.3	7.1.4	Supported with RHEL 7.7 IBM Spectrum Scale versions <sup>1</sup> : 5.1.0.1 - 5.1.0.3

<sup>1</sup> See [IBM Spectrum Scale Big Data and Analytics Support Matrix](#) for the latest support information.

# Data Migration Versus Upgrade

Recommendations on whether to upgrade to Cloudera Base on premises or migrate workloads to Cloudera on cloud.

- Data migration refers to moving existing CDH or HDP workloads to Cloudera on cloud or to a new installation of Cloudera Base on premises.
- Upgrade refers to a full in-place upgrade of CDH or HDP to Cloudera Base on premises.

The path to Cloudera that works best for you depends on the size of your clusters, the types of workloads you are running, and whether you want to move workloads to the Cloud, stay exclusively on-prem, or use a combination of on-prem and cloud.



**Note:** These recommendations apply to both CDH and HDP clusters.

## On-prem cluster less than 50 hosts with Hive or Impala

If you are running Hive or Impala workloads without HBase on an on-prem cluster with less than 50 hosts, and less than 5 services running on the cluster:

- Migrate workloads to Cloudera Data Warehouse on Cloudera.

## On-prem cluster less than 50 hosts with HBase

If you are running HBase workloads without Hive or Impala on an on-prem cluster with less than 50 hosts, and less than 5 services running on the cluster:

- Migrate workloads to Cloudera Data Hub on Cloudera on cloud and use the Cloudera Operational Database cluster template.

## On-prem cluster less than 50 hosts with Spark

If you are running Spark workloads without Kafka, NiFi, or Storm on an on-prem cluster with less than 50 hosts, and less than 5 services running on the cluster:

- Migrate workloads to Cloudera Data Hub on Cloudera on cloud and use the Cloudera Data Engineering cluster template.

## On-prem cluster more than 800 hosts

If you are running workloads on an on-prem cluster with more than 800 hosts:

- Split the cluster up into multiple 100-300 node clusters and upgrade to Cloudera Base on premises.

## Multiple on-prem clusters with more than 100 hosts

If you are running workloads on multiple on-prem clusters with a combined total of more than 100 hosts, and less than 50 services in total:

- Consolidate the clusters into one 100-300 node cluster and upgrade to Cloudera Base on premises.