

# Upgrading to the Latest Version of Cloudera Data Science Workbench on CDH

Date published: 2020-02-28

Date modified: 2020-12-15



# Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

# Contents

<b>Upgrading to the Latest Version of Cloudera Data Science Workbench on CDH.....</b>	<b>4</b>
Upgrading Cloudera Data Science Workbench 1.7.2 or higher from CDH 6 to CDP Private Cloud Base 7.x.....	4
Upgrading Cloudera Data Science Workbench 1.9.0 Using Cloudera Manager.....	5
Upgrading CSD Deployments from CDH 5 to CDH 6.....	8
Upgrading RPM Deployments from CDH 5 to CDH 6.....	9
Migrating from an RPM-based Deployment to the Latest 1.9.0 CSD.....	10
Upgrading Cloudera Data Science Workbench Using Packages.....	14

# Upgrading to the Latest Version of Cloudera Data Science Workbench on CDH

This topic walks you through the upgrade paths available for Cloudera Data Science Workbench.



**Note:** The expected host IP of CDSW\_PUBLIC\_PORT has been changed from 0.0.0.0 to localhost (127.0.0.1) as of CDSW 1.8.x. This will affect the ability of analytical applications to connect for users who are not authenticated by CDSW. Existing applications listening on "0.0.0.0:CDSW\_PUBLIC\_PORT" must be migrated to "localhost:CDSW\_PUBLIC\_PORT".

Depending on your existing deployment, choose from one of the upgrade paths listed in the following table.

Upgrading from CDH 5 CDH 6 - If you are currently running Cloudera Data Science Workbench on CDH 5 and want to upgrade your cluster to CDH 6, also see [Upgrading Cloudera Data Science Workbench from CDH 5 to CDH6..](#)

Upgrade Path	Link to Instructions
CSD CSD Upgrading from an existing CSD-based deployment to the latest CSD and parcel.	<a href="#">Upgrading Cloudera Data Science Workbench 1.9.0 Using Cloudera Manager</a>
RPM CSD Migrating from an RPM-based deployment to the latest CSD and parcel-based deployment.	<a href="#">Migrating from an RPM-based Deployment to the Latest 1.9.0 CSD</a>
RPM RPM Upgrading an existing RPM-based deployment to the latest RPM.	<a href="#">Upgrading Cloudera Data Science Workbench 1.9.0 Using Packages</a> Note that you cannot use Cloudera Manager for this upgrade path.

## Upgrading Cloudera Data Science Workbench 1.7.2 or higher from CDH 6 to CDP Private Cloud Base 7.x

The following general outline describes how to upgrade your Cloudera Data Science Workbench cluster from CDH 6 to CDP Private Cloud Base. Refer to the relevant linked Cloudera Manager and CDH upgrade documentation for detailed steps required for this procedure.

### Procedure

- Before upgrading to 1.8, if you are using LDAPs, make sure the certificate uses a key of at least 2048 bits.  
If CDSW is configured to talk to an LDAP server for authentication, and that LDAP connection is over https, and the certificate uses a 1024 bit RSA key, connection attempts will fail and no one will be able to log in.
- Migrate CDSW from RPM to CSD: see the instructions to [migrate to CDP files](#).
- Download the correct CDSW CSD files for CDP Private Cloud Base.  
The correct CSD file names are:
  - For CDH 6: CLOUDERA\_DATA\_SCIENCE\_WORKBENCH\_CDH6\_1.x.y.jar
  - For CDP Private Cloud Base: CLOUDERA\_DATA\_SCIENCE\_WORKBENCH\_CDPDC\_1.x.y.jar

4. Upgrade cluster from CDH 6 to CDP Private Cloud Base 7.x:
  - a) Stop Cloudera Data Science Workbench.
  - b) Log on to the Cloudera Manager Server host.
  - c) Place the CSD files in /opt/cloudera/csd.  
This is the default location for CSD files.
  - d) Restart Cloudera Manager Server.
  - e) [Upgrade to Cloudera Data Science Workbench 1.7.2 \(or higher\)](#). During the upgrade process, as you install, distribute, and activate the new parcel, take care to ensure that both CDSW CSD files (for CDH 6 and CDP Private Cloud Base) are present on the Cloudera Manager Server host.
  - f) Upgrade CDP Private Cloud Base based on the CDP Private Cloud Base [upgrade documentation](#).
  - g) Remove any existing CDH 6 CSDs from the Cloudera Manager Server host. This step is optional.
  - h) Restart Cloudera Data Science Workbench.

## Upgrading Cloudera Data Science Workbench 1.9.0 Using Cloudera Manager

This topic describes how to upgrade a CSD and parcel-based deployment to the latest version of Cloudera Data Science Workbench 1.9.0.

### Before you begin



**Note:** Cloudera Data Science Workbench only supports upgrades to version 1.9.0 from the last two versions (1.8.x, 1.7.x). If you are using a lower version, you must first upgrade to version 1.7.x or 1.8.x and then upgrade to version 1.9.0.

### About this task

For installation and upgrades, you must manually add the Remote Parcel Repository URL for your CDSW version to Cloudera Manager.

### Procedure

1. Before you begin the upgrade process, make sure you read the Cloudera Data Science Workbench Release Notes relevant to the version you are upgrading to/from.
2. Stop the Cloudera Data Science Workbench service in Cloudera Manager.
3. If the patches folder exists, remove all of patches and the /patches/ folder.  
You can reapply patches if they are needed for the current release.
4. (Strongly Recommended) On the master host, backup all your application data that is stored in the /var/lib/cdsw directory.

To create the backup, run the following command on the master host:

```
tar cvzf cdsw.tar.gz /var/lib/cdsw/*
```

5. Deactivate the existing Cloudera Data Science Workbench parcel. Go to the Cloudera Manager Admin Console. In the top navigation bar, click **Hosts** **Parcels** .  
Locate the current active CDSW parcel and click **Deactivate**. On the confirmation pop-up, select **Deactivate Only** and click **OK**.

**6. Download and save the latest Cloudera Data Science Workbench CSD to the Cloudera Manager Server host.**

**Note:** To download Cloudera Data Science Workbench, you must have an active subscription agreement along with the required authentication credentials (namely, the username and password). The authentication credentials are provided in an email sent to the customer account from Cloudera when a new license is issued.

If you do not have the authentication credentials, contact your account representative to receive the same.

You can download the CDSW version 1.9 or later using one of the following methods:

- Log in to the Cloudera Downloads web page and download the required files.
- Use the URLs listed in this section and provide the login information as provided with your Cloudera subscription.

- a) Download the Cloudera Data Science Workbench CSD. Make sure you download the CSD that corresponds to the version of CDH or Cloudera Runtime you are using.

- CDP Data Center

```
https://archive.cloudera.com/p/cdsw1/1.9.0/csd/CLOUDERA_DATA_SCIENCE  
_WORKBENCH-CDPDC-1.9.0.jar
```

OR

- CDH 6

```
https://archive.cloudera.com/p/cdsw1/1.9.0/csd/CLOUDERA_DATA_SCIENCE  
_WORKBENCH-CDH6-1.9.0.jar
```

OR

- CDH 5

```
https://archive.cloudera.com/p/cdsw1/1.9.0/csd/CLOUDERA_DATA_SCIENCE  
_WORKBENCH-CDH5-1.9.0.jar
```

- b) Log on to the Cloudera Manager Server host, and place the CSD file under `/opt/cloudera/csd`, which is the default location for CSD files.
- c) Delete any CSD files belonging to older versions of Cloudera Data Science Workbench from `/opt/cloudera/csd`.

This is required because older versions of the CSD will not work with the latest Cloudera Data Science Workbench parcel. Make sure your CSD and parcel are always the same version.

After you delete the file(s) belonging to the older version, you should still have two files for the current version; one for CDH5 (for example, `CDSW1.9-CDH5.jar`) and one for CDH6 (for example, `CDSW1.9-CDH6.jar`). You'll need both of these files to upgrade.

Note: If you have previously configured a custom location for CSD files, place the CSD file there, and delete any CSDs belonging to older versions of Cloudera Data Science Workbench. For help, refer the Cloudera Manager documentation at [Configuring the Location of Custom Service Descriptor Files](#).

- d) Set the CSD file ownership to `cloudera-scm:cloudera-scm` with permission 644.
- e) Restart the Cloudera Manager Server:

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

- f) Log into the Cloudera Manager Admin Console and restart the Cloudera Management Service.
1. Select Clusters > Cloudera Management Service.
  2. Select Actions > Restart.

**7. Distribute and activate the new parcel on your cluster.**

- a) Log into the Cloudera Manager Admin Console.
- b) Click **Hosts Parcels** in the main navigation bar.
- c) Add the Cloudera Data Science Workbench parcel repository URL to Cloudera Manager.
  1. On the **Parcels** page, click **Configuration**.
  2. In the **Remote Parcel Repository URLs** list, click the addition symbol to create a new row.
  3. Enter the path to the repository.

**Cloudera Data Science Workbench 1.9.0**

```
https://archive.cloudera.com/p/cdsw1/1.9.0/parcels/
```

4. Click **Save Changes**.
  - d) Go back to the **Hosts Parcels** page. The latest parcel should now appear in the set of parcels available for download. Click **Download**. Once the download is complete, click **Distribute** to distribute the parcel to all the CDH hosts in your cluster. Then click **Activate**. For more detailed information on each of these tasks, see [Managing Parcels](#).
- 8. Run the Prepare Node command on all Cloudera Data Science Workbench hosts.**


- a) Before you run **Prepare Node**, you must make sure that the command is allowed to install all the required packages on your cluster. This is controlled by the **Install Required Packages** property.
  1. Navigate to the CDSW service.
  2. Click **Configuration**.
  3. Search for the **Install Required Packages** property. If this property is enabled, you can move on to the next step and run **Prepare Node**.

However, if the property has been disabled, you must either enable it or manually install the following packages on all Cloudera Data Science Workbench gateway hosts.

```
nfs-utils
libseccomp
lvm2
bridge-utils
libtool-ltdl
iptables
rsync
policycoreutils-python
selinux-policy-base
selinux-policy-targeted
ntp
ebtables
bind-utils
nmap-ncat
openssl
e2fsprogs
redhat-lsb-core
conntrack-tools
socat
```

- b) Run the **Prepare Node** command.
  1. In Cloudera Manager, navigate to the Cloudera Data Science Workbench service.
  2. Click the **Instances** tab.
  3. Use the checkboxes to select all host instances and click **Actions for Selected (x)**.
  4. Click **Prepare Node**. Once again, click **Prepare Node** to confirm the action.

9. Log into the Cloudera Manager Admin Console and restart the Cloudera Data Science Workbench service.

- a) On the Home Status tab, click  to the right of the CDSW service and select Restart from the dropdown.
- b) Confirm your choice on the next screen. Note that a complete restart of the service will take time. Even though the CDSW service status shows Good Health, the application itself will take some more time to get ready.

10. Upgrade Projects to Use the Latest Base Engine Images

If the release you have just upgraded to includes a new version of the base engine image, you will need to manually configure existing projects to use the new engine. Cloudera recommends you do so to take advantage of any new features and bug fixes included in the newly released engine. For example:

- Container Security

Security best practices dictate that engine containers should not run as the root user. Engines (v7 and lower) briefly initialize as the root user and then run as the cdsw user. Engines v8 (and higher) now follow the best practice and run only as the cdsw user. For more details, see [Allow containers to run as root](#).

- CDH 6 Compatibility

The base engine image you use must be compatible with the version of CDH you are running. This is especially important if you are running workloads on Spark. Older base engines (v6 and lower) cannot support the latest versions of CDH 6. If you want to run Spark workloads on CDH 6, you must upgrade your projects to base engine 7 (or higher).

- Editors

Engines v8 (and higher) ships with the browser-based IDE, Jupyter, preconfigured and can be selected from the Start Session menu.

To upgrade a project to the new engine, go to the project's Settings Engine page and select the new engine from the dropdown. If any of your projects are using custom extended engines, you will need to modify them to use the new base engine image.

11. (GPU-enabled Deployments) Remove nvidia-docker1 and Upgrade NVIDIA Drivers to 410.xx or higher

Perform the following steps to make sure you can continue to leverage GPUs for workloads on Cloudera Data Science Workbench 1.6 (and higher).

- a) Remove nvidia-docker1. Cloudera Data Science Workbench (version 1.6 and higher) ships with nvidia-docker2 installed by default.

```
yum remove nvidia-docker
```

Perform this step on all hosts that have GPUs attached to them.

- b) Upgrade your NVIDIA driver to version 410.xx (or higher). This must be done because nvidia-docker2 does not support lower versions of NVIDIA drivers.

1. Stop Cloudera Data Science Workbench.

Depending on your deployment, either stop the CDSW service in Cloudera Manager (for CSDs) or run `cdsw stop` on the Master host (for RPMs).

2. Reboot the GPU-enabled hosts.

3. Install a supported version of the NVIDIA driver (410.xx or higher) on all GPU-enabled hosts.

4. Start Cloudera Data Science Workbench.

Depending on your deployment, either start the CDSW service in Cloudera Manager (for CSDs) or run `cdsw start` on the Master host (for RPMs).

## Upgrading CSD Deployments from CDH 5 to CDH 6

This section provides a general outline on how to go about upgrading your Cloudera Data Science Workbench cluster from CDH 5 to CDH 6. Refer the relevant linked Cloudera Manager and CDH upgrade documentation for the detailed steps required for this procedure.



### About this task

Upgrading CSD Deployments from CDH 5 to CDH 6

Starting with version 1.5, Cloudera Data Science Workbench publishes two separate CSD files: one for CDH 5 and one for CDH 6. Check the CSD file name to ensure that you are using the correct CSD file for your cluster. For example:

- CDH 6 - CLOUDERA\_DATA\_SCIENCE\_WORKBENCH\_CDH6\_1.x.y.jar
- CDH 5 - CLOUDERA\_DATA\_SCIENCE\_WORKBENCH\_CDH5\_1.x.y.jar

Use the following path to upgrade from running a CSD-based Cloudera Data Science Workbench deployment on CDH 5 to running on CDH 6:

### Procedure

1. Upgrade to Cloudera 6.1 (or higher).
2. Stop Cloudera Data Science Workbench.
3. Remove both Spark gateway roles from all CDSW hosts.
4. Delete the /etc/spark and /etc/spark2 directories.
5. Download both of the CDSW Cloudera Data Science Workbench CSD files for the latest version. For example:
  - CDSW1.8-CDH6.jar
  - CDSW1.8-CDH5.jar

At this point, you should have three CSV files. One original file (for example, CDSW1.5-CDH5.jar) and two new files (for example, CDSW1.8-CDH6.jar and CDSW1.8-CDH5.jar).

6. Log on to the Cloudera Manager Server host, and place the new CDSW files under /opt/cloudera/csd, which is the default location for CSDs.
7. Restart the Cloudera Manager Server.
8. [Upgrade to Cloudera Data Science Workbench 1.5 \(or higher\)](#). During the upgrade process, as you install, distribute, and activate the new parcel, take care to ensure that both the CDSW CSDs (for CDH 5 and CDH 6) are present on the Cloudera Manager Server host.
9. [Use the Cloudera Manager Upgrade Wizard to upgrade from CDH 5 to CDH 6.1 \(or higher\)](#).

As part of the upgrade, the wizard will also remove the Spark 2 parcel from all your cluster hosts. With CDH 6, Spark 2 ships as a part of CDH. The add-on parcel is no longer required.

Cloudera Manager 6 can differentiate between the two active CSDs and will select the right one based on the version of CDH running. Because you already have the CDH 6-compatible CSD installed, no further steps are needed.
10. (Optional) Remove any existing CDH 5 CSDs from the Cloudera Manager Server host.
11. Add the Spark gateway back in.
12. Redeploy the client configurations.
13. Restart Cloudera Data Science Workbench.

## Upgrading RPM Deployments from CDH 5 to CDH 6

Cloudera Data Science Workbench ships a single RPM that can be used to install CDSW on both, CDH 5, and CDH 6 clusters.

### About this task

The upgrade path for RPM deployments is:

### Procedure

1. Upgrade to Cloudera Manager 6.1 (or higher).

2. Use the Cloudera Manager Upgrade Wizard to upgrade from CDH 5 to CDH 6 (or higher). As part of the upgrade, the wizard will also remove the Spark 2 parcel from all your cluster hosts. This is because with CDH 6, Spark 2 ships as a part of CDH. The add-on parcel is no longer required.
3. Upgrade to the latest Cloudera Data Science Workbench RPM.

## Migrating from an RPM-based Deployment to the Latest 1.9.0 CSD

This topic describes how to migrate from an RPM-based deployment to the latest 1.9.0 CSD and parcel-based deployment.

### Before you begin

Make sure you read the [Deploying Cloudera Data Science Release Notes](#) relevant to the version you are migrating to/from.

### Procedure

1. Save a backup of the Cloudera Data Science Workbench configuration file located at `/etc/cdsd/config/cdsd.conf`.
2. Stop the Cloudera Data Science Workbench service in Cloudera Manager.
3. Stop the Cloudera Data Science Workbench service in Cloudera Manager.
  - a) Delete the 2 patch files: `/etc/cdsd/patches/default/deployment/ingress-controller.yaml` and `/etc/cdsd/patches/default/deployment/tcp-ingress-controller.yaml`.
  - b) Delete every empty folder from the `/etc/cdsd/patches` directory.
  - c) Delete the `/etc/cdsd/patches` directory if it is empty.
4. (Strongly Recommended) On the master host, backup all your application data that is stored in the `/var/lib/cdsd` directory.

To create the backup, run the following command on the master host:

```
tar cvzf cdsd.tar.gz /var/lib/cdsd/*
```

5. Save a backup of the Cloudera Data Science workbench configuration file at:

```
/etc/cdsd/config/cdsd.conf
```


6. Uninstall the previous release of Cloudera Data Science Workbench. Perform this step on the master host, as well as all the worker hosts.

```
yum remove cloudera-data-science-workbench
```

7. Install the latest version of Cloudera Data Science Workbench using the CSD and parcel. Note that when you are configuring role assignments for the Cloudera Data Science Workbench service, the Master role must be assigned to the same host that was running as master prior to the upgrade.

For installation instructions, see [Installing Cloudera Data Science Workbench 1.9.0 Using Cloudera Manager](#). You might be able to skip the first few steps assuming you have the wildcard DNS domain and block devices already set up.

8. Use your copy of the backup `cdsw.conf` created in Step 3 to recreate those settings in Cloudera Manager by configuring the corresponding properties under the Cloudera Data Science Workbench service.
  - a) Log into the Cloudera Manager Admin Console.
  - b) Go to the Cloudera Data Science Workbench service.
  - c) Click the Configuration tab.
  - d) The following table lists all the `cdsw.conf` properties and their corresponding Cloudera Manager properties (in bold). Use the search box to bring up the properties you want to modify.
  - e) Click Save Changes.

cdsw.conf Property	Corresponding Cloudera Manager Property and Description
TLS_ENABLE	<p>Enable TLS: Enable and enforce HTTPS (TLS/SSL) access to the web application (optional). Both internal and external termination are supported. To enable internal termination, you must also set the TLS Certificate for Internal Termination and TLS Key for Internal Termination parameters. If these parameters are not set, terminate TLS using an external proxy.</p> <p>For more details on TLS termination, see <a href="#">Enabling TLS/SSL for Cloudera Data Science Workbench</a>.</p>
TLS_CERT TLS_KEY	<p>TLS Certificate for Internal Termination, TLS Key for Internal Termination</p> <p>Complete path to the certificate and private key (in PEM format) to be used for internal TLS termination. Set these parameters only if you are not terminating TLS externally. You must also set the Enable TLS property to enable and enforce termination. The certificate must include both DOMAIN and *.DOMAIN as hostnames.</p> <p>Self-signed certificates are not supported unless trusted fully by clients. Accepting an invalid certificate manually can cause connection failures for unknown subdomains. Set these only if you are not terminating TLS externally. For details on certificate requirements and enabling TLS termination, see <a href="#">Enabling TLS/SSL for Cloudera Data Science Workbench</a>.</p>
TLS_ROOTCA	<p> <b>Note:</b> This property is not available in Cloudera Manager. It must be set in the Cloudera Data Science Workbench Site Administration panel after installation is complete. For instructions, see <a href="#">Configuring Custom Root CA Certificate</a>.</p> <p>If your organization uses an internal custom Certificate Authority, you can use this field to paste in the contents of your internal CA's root certificate file.</p> <p>The contents of this field are then inserted into the engine's root certificate store every time a session (or any workload) is launched. This allows processes inside the engine to communicate securely with the ingress controller.</p>

cdsw.conf Property	Corresponding Cloudera Manager Property and Description
HTTP_PROXY HTTPS_PROXY	<p>HTTP Proxy, HTTPS Proxy</p> <p>If your deployment is behind an HTTP or HTTPS proxy, set the respective HTTP Proxy or HTTPS Proxy property to the hostname of the proxy you are using.</p> <pre>http://&lt;proxy_host&gt;:&lt;proxy-port&gt;</pre> <p>or</p> <pre>https://&lt;proxy_host&gt;:&lt;proxy-port&gt;</pre> <p>If you are using an intermediate proxy such as <a href="#">Cntlm</a> to handle NTLM authentication, add the Cntlm proxy address to the HTTP Proxy or HTTPS Proxy fields. That is, either http://localhost:3128 or https://localhost:3128 respectively.</p> <p>If the proxy server uses TLS encryption to handle connection requests, you will need to add the proxy's root CA certificate to your host's store of trusted certificates. This is because proxy servers typically sign their server certificate with their own root certificate. Therefore, any connection attempts will fail until the Cloudera Data Science Workbench host trusts the proxy's root CA certificate. If you do not have access to your proxy's root certificate, contact your Network / IT administrator.</p> <p>To enable trust, copy the proxy's root certificate to the trusted CA certificate store (ca-trust) on the Cloudera Data Science Workbench host.</p> <pre>cp /tmp/&lt;proxy-root-certificate&gt;.crt /etc/pki/ca-trust/ source/anchors/</pre> <p>Use the following command to rebuild the trusted certificate store.</p> <pre>update-ca-trust extract</pre>
ALL_PROXY	SOCKS Proxy: If a SOCKS proxy is in use, set this parameter to socks5://<host>:<port>/.

cdsw.conf Property	Corresponding Cloudera Manager Property and Description
NO_PROXY	<p>No Proxy: Comma-separated list of hostnames that should be skipped from the proxy.</p> <p>Starting with version 1.4, if you have defined a proxy in the HTTP_PROXY(S) or ALL_PROXY properties, Cloudera Data Science Workbench automatically appends the following list of IP addresses to the NO_PROXY configuration. Note that this is the minimum required configuration for this field.</p> <p>This list includes 127.0.0.1, localhost, and any private Docker registries and HTTP services inside the firewall that Cloudera Data Science Workbench users might want to access from the engines.</p> <pre>"127.0.0.1,localhost,100.66.0.1,100.66.0.2,100.66.0.3,100.66.0.4,100.66.0.5,100.66.0.6,100.66.0.7,100.66.0.8,100.66.0.9,100.66.0.10,100.66.0.11,100.66.0.12,100.66.0.13,100.66.0.14,100.66.0.15,100.66.0.16,100.66.0.17,100.66.0.18,100.66.0.19,100.66.0.20,100.66.0.21,100.66.0.22,100.66.0.23,100.66.0.24,100.66.0.25,100.66.0.26,100.66.0.27,100.66.0.28,100.66.0.29,100.66.0.30,100.66.0.31,100.66.0.32,100.66.0.33,100.66.0.34,100.66.0.35,100.66.0.36,100.66.0.37,100.66.0.38,100.66.0.39,100.66.0.40,100.66.0.41,100.66.0.42,100.66.0.43,100.66.0.44,100.66.0.45,100.66.0.46,100.66.0.47,100.66.0.48,100.66.0.49,100.66.0.50,100.77.0.10,100.77.0.128,100.77.0.129,100.77.0.130,100.77.0.131,100.77.0.132,100.77.0.133,100.77.0.134,100.77.0.135,100.77.0.136,100.77.0.137,100.77.0.138,100.77.0.139"</pre>
NVIDIA_GPU_ENABLED	<p>Enable GPU Support: When this property is enabled, GPUs installed on Cloudera Data Science Workbench hosts will be available for use in its workloads. By default, this parameter is disabled.</p> <p>For instructions on how to enable GPU-based workloads on Cloudera Data Science Workbench, see <a href="#">Configuring Custom Root CA Certificate</a>.</p>

9. Cloudera Manager will prompt you to restart the service if needed.

10. If the release you have just upgraded to includes a new version of the base engine image (see release notes), you will need to manually configure existing projects to use the new engine. Cloudera recommends you do so to take advantage of any new features and bug fixes included in the newly released engine. For example:

- Container Security

Security best practices dictate that engine containers should not run as the root user. Engines (v7 and lower) briefly initialize as the root user and then run as the cdsw user. Engines v8 (and higher) now follow the best practice and run only as the cdsw user. For more details, see [Restricting User-Created Pods](#).

- CDH 6 Compatibility

The base engine image you use must be compatible with the version of CDH you are running. This is especially important if you are running workloads on Spark. Older base engines (v6 and lower) cannot support

the latest versions of CDH 6. If you want to run Spark workloads on CDH 6, you must upgrade your projects to base engine 7 (or higher).

- Editors

Engines v8 (and higher) ships with the browser-based IDE, Jupyter, preconfigured and can be selected from the Start Session menu.

To upgrade a project to the new engine, go to the project's Settings > Engine page and select the new engine from the dropdown. If any of your projects are using custom extended engines, you will need to modify them to use the new base engine image.

#### 11. (GPU-enabled Deployments) Remove nvidia-docker1 and Upgrade NVIDIA Drivers to 410.xx or higher

Perform the following steps to make sure you can continue to leverage GPUs for workloads on Cloudera Data Science Workbench 1.6 (and higher).

- a) Remove nvidia-docker1. Cloudera Data Science Workbench (version 1.6 and higher) ships with nvidia-docker2 installed by default.

Perform this step on all hosts that have GPUs attached to them.

- b) Upgrade your NVIDIA driver to version 410.xx (or higher). This must be done because nvidia-docker2 does not support lower versions of NVIDIA drivers.

- Stop Cloudera Data Science Workbench.

Depending on your deployment, either stop the CDSW service in Cloudera Manager (for CSDs) or run `cdsw stop` on the Master host (for RPMs).

- Reboot the GPU-enabled hosts. Install a supported version of the NVIDIA driver (410.xx or higher) on all GPU-enabled hosts.
- Start Cloudera Data Science Workbench.

Depending on your deployment, either start the CDSW service in Cloudera Manager (for CSDs) or run `cdsw start` on the Master host (for RPMs).

## Upgrading Cloudera Data Science Workbench Using Packages

This topic describes how to upgrade an RPM-based deployment to the latest version of Cloudera Data Science Workbench 1.9.0.

### Before you begin

Before you start upgrading Cloudera Data Science Workbench, read the Cloudera Data Science Workbench Release Notes relevant to the version you are upgrading to.

### Procedure

1. Run the following command on all Cloudera Data Science Workbench hosts (master and workers) to stop Cloudera Data Science Workbench.

```
cdsw stop
```

2. (Upgrading from CDSW 1.7.1 with patch) Perform this step only if you are upgrading from CDSW 1.7.1 with an applied patch.
  - a) Delete the 2 patch files: `/etc/cdsw/patches/default/deployment/ingress-controller.yaml` and `/etc/cdsw/patches/default/deployment/tcp-ingress-controller.yaml`.
  - b) Delete every empty folder from the `/etc/cdsw/patches` directory.
  - c) Delete the `/etc/cdsw/patches` directory if it is empty.

3. (Strongly Recommended) On the master host, backup all your application data that is stored in the `/var/lib/cdsw` directory.

To create the backup, run the following command on the master host:

```
tar cvzf cdsw.tar.gz /var/lib/cdsw/*
```

4. Save a backup of the Cloudera Data Science workbench configuration file at:

```
/etc/cdsw/config/cdsw.conf
```

5. Uninstall the previous release of Cloudera Data Science Workbench. Perform this step on the master host, as well as all the worker hosts.

```
yum remove cloudera-data-science-workbench
```

6. Install the latest version of Cloudera Data Science Workbench on the master host and on all the worker hosts. During the installation process, you might need to resolve certain incompatibilities in `cdsw.conf`. Even though you will be installing the latest RPM, your previous configuration settings in `cdsw.conf` will remain unchanged. Depending on the release you are upgrading from, you will need to modify `cdsw.conf` to ensure it passes the validation checks run by the release.

To install the latest version of Cloudera Data Science Workbench, follow the same process to install the package as you would for a fresh installation.

- [Install Cloudera Data Science Workbench on the Master Host](#)
- (Optional) [Install Cloudera Data Science Workbench on Worker Hosts](#)

7. Upgrade Projects to Use the Latest Base Engine Images

If the release you have just upgraded to includes a new version of the base engine image, you will need to manually configure existing projects to use the new engine. Cloudera recommends you do so to take advantage of any new features and bug fixes included in the newly released engine. For example:

- Container Security

Security best practices dictate that engine containers should not run as the root user. Engines (v7 and lower) briefly initialize as the root user and then run as the `cdsw` user. Engines v8 (and higher) now follow the best practice and run only as the `cdsw` user. For more details, see [Restricting User-Created Pods](#).

- CDH 6 Compatibility

The base engine image you use must be compatible with the version of CDH you are running. This is especially important if you are running workloads on Spark. Older base engines (v6 and lower) cannot support the latest versions of CDH 6. If you want to run Spark workloads on CDH 6, you must upgrade your projects to base engine 7 (or higher).

- Editors

Engines v8 (and higher) ships with the browser-based IDE, Jupyter, preconfigured and can be selected from the Start Session menu.

To upgrade a project to the new engine, go to the project's `Settings Engine` page and select the new engine from the dropdown. If any of your projects are using custom extended engines, you will need to modify them to use the new base engine image.

**8. (GPU-enabled Deployments) Remove nvidia-docker1 and Upgrade NVIDIA Drivers to 410.xx or higher**

Perform the following steps to make sure you can continue to leverage GPUs for workloads on Cloudera Data Science Workbench 1.6 (and higher).

- a) Remove nvidia-docker1. Cloudera Data Science Workbench (version 1.6 and higher) ships with nvidia-docker2 installed by default.

```
yum remove nvidia-docker
```

Perform this step on all hosts that have GPUs attached to them.

- b) Upgrade your NVIDIA driver to version 410.xx (or higher). This must be done because nvidia-docker2 does not support lower versions of NVIDIA drivers.

- Stop Cloudera Data Science Workbench.

Depending on your deployment, either stop the CDSW service in Cloudera Manager (for CSDs) or run `cdsw stop` on the Master host (for RPMs).

- Reboot the GPU-enabled hosts.
- Install a supported version of the NVIDIA driver (410.xx or higher) on all GPU-enabled hosts.
- Start Cloudera Data Science Workbench.

Depending on your deployment, either start the CDSW service in Cloudera Manager (for CSDs) or run `cdsw start` on the Master host (for RPMs).