

Cloudera AI 1.5.4

## CDSW to Cloudera AI migration

Date published: 2022-12-14

Date modified: 2025-03-18

# CLOUDERA

<https://docs.cloudera.com/>

# Legal Notice

© Cloudera Inc. 2025. 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>Migrating Data Science Workbench (CDSW) to Cloudera AI.....</b>	<b>4</b>
Prerequisites for CDSW to Cloudera AI migration.....	4
Repurposing CDSW nodes for Cloudera AI.....	6
Removing cluster hosts and Kubernetes nodes.....	7
Customizing CDSW for migrating host mounts.....	13
Using the CDSW to Cloudera AI Migration tool.....	13
Troubleshooting preflight migration check issues.....	24
Limitations of migrating CDSW to Cloudera AI.....	25
Post-migration tasks.....	26
Troubleshooting CDSW migration to Cloudera AI.....	27

# Migrating Data Science Workbench (CDSW) to Cloudera AI

If you use Cloudera Data Science Workbench (CDSW) 1.10.0 and higher version on premises, consider migrating to Cloudera AI as soon as possible. Cloudera recommends performing this migration because CDSW end-of-support is imminent.

Before you begin the migration tasks, you can optionally repurpose your CDSW hosts instead of adding hardware for installing Cloudera AI on premises. Next, you install Cloudera AI on the hosts. Finally, the UI-driven migration tool runs scripts in the background to migrate your workload automatically after installing Cloudera AI on the cluster with your deployed CDSW.

## Prerequisites for CDSW to Cloudera AI migration

Before migrating from Cloudera Data Science Workbench (CDSW) to Cloudera AI in Cloudera on premises, you must meet a number of prerequisites to succeed. A prerequisite for migration is the installation of Cloudera AI on your Cloudera Data Science Workbench base cluster.

### About this task

The following table presents the supported migration version combinations for Cloudera Data Science Workbench and Cloudera AI:

**Table 1: Supported migration versions for Cloudera Data Science Workbench and Cloudera AI**

Supported CDSW versions	Target Cloudera AI on premises version
CDSW 1.10.0- 1.10.4	1.5.3 1.5.4 SP1 (recommended version) Upgrade to CDSW 1.10.5 before migrating to Cloudera AI.
CDSW 1.10.5	1.5.3 1.5.4 1.5.4 SP1 (recommended version)



### Note:

With CDSW 1.10.3 or 1.10.4, an upgrade to CDSW 1.10.5 is recommended and a migration to Cloudera AI on premises version higher than 1.5.2.

Migration from CDSW, configured with LDAP, SAML, or LOCAL authentication to Cloudera AI, is supported, but the automatic migration is supported only if CDSW is running with LDAP. The migration process does not automatically migrate your authentication configurations. Therefore, setting up LDAP in CDSW prior to migration is part of the migration procedure.

The migration does not migrate your CDSW endpoint connections. Therefore, post-migration instructions include setting up LDAP, endpoint connections, and DNS on Cloudera AI, as well as downloading CDSW-related Grafana dashboards, so you can upload them after migration to Cloudera AI.

## Procedure

1. You must have a CDSW 1.10.0 or higher version cluster in Cloudera; otherwise, choose one of the following options:
  - If you have a CDSW installation in either CDH or HDP, migrate to on premises 1.5.1 or higher version, and then migrate CDSW to Cloudera AI.
  - If you have CDSW installation earlier than 1.10.0, upgrade to CDSW 1.10.0 or higher versions.
2. If you do not have LDAP set up in your CDSW cluster on Cloudera, set up LDAP before pre-migration tasks. For guidelines on setting up LDAP, see [Configuring External Authentication with LDAP and SAML](#).  
The migration process cannot succeed without authentication.
3. Meet the Cloudera AI software requirements for on premises, including storage, for installing Cloudera AI on Cloudera on premises 1.5.1 or higher version. For Cloudera AI software requirements for on premises, see [Cloudera AI software requirements for Cloudera](#).
4. Backup CDSW data. For details on how to backup CDSW data, see [Backup and Disaster Recovery for Cloudera Data Science Workbench](#).
5. In CDSW, export your Grafana dashboards. For details on how to export Grafana dashboards, see [Export and import | Grafana documentation](#).
6. Note the connections of endpoints in your CDSW cluster, consider your custom settings.  
You need to use this information after migration to set up endpoints in your on premises cluster.
7. If you customized your DNS configuration, make notes your custom settings to be able to customize your DNS configuration after migration.  
If you did not customize your DNS configuration, the migration tool configures DNS in your on premises cluster.
8. Gather information about your LDAP configurations on CDSW.  
After migration, you must set up LDAP again on the Cloudera AI cluster. The LDAP configuration is not migrated.
9. In CDSW, manually back up the custom DNS configuration for Kube-DNS, and then migrate your custom configuration to Cloudera AI.  
Cloudera AI uses the core-DNS, which is incompatible with the CDSW Kube-DNS.
10. In Cloudera Manager, select install and upgrade to Cloudera on premises 1.5.1 or higher version using the Cloudera Embedded Container Service on your CDSW cluster.  
Migration of your CDSW workloads to Cloudera AI on OpenShift is not supported.
11. During the installation of Cloudera Data Services on premises using Cloudera Embedded Container Service set up a network connection between CDSW and the Cloudera on premises cluster if you select Airgap.
12. Enable those Cloudera AI features during installation that you were using in CDSW.  
For example, enable model metrics and monitoring.

Production Cloudera AI

Enable Governance ⓘ

Enable Model Metrics ⓘ

Other Settings

Enable TLS ⓘ

Enable Monitoring ⓘ

If you do not enable the same, or similar, Cloudera AI features during installation that you were using in CDSW, you will not be able to use the Cloudera AI features.

## Related Information

[Configuring External Authentication with LDAP and SAML](#)

[CDP Upgrade and Migrations Paths](#)

[CML software requirements for Private Cloud](#)

[Backup and Disaster Recovery for Cloudera Data Science Workbench](#)

[Export and import | Grafana documentation](#)

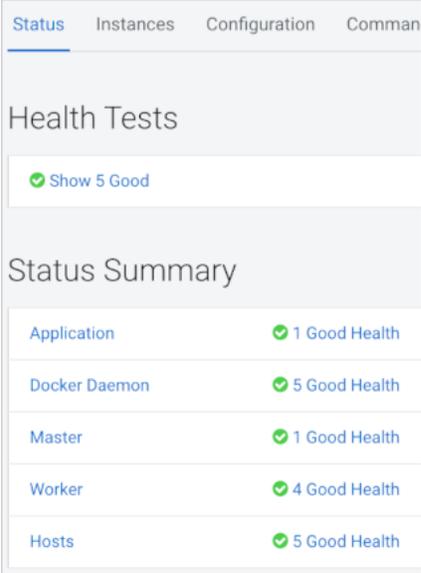
[CML software requirements for Private Cloud](#)

[CML software requirements for Private Cloud](#)

## Repurposing CDSW nodes for Cloudera AI

The description of an example scenario prepares you to repurpose your existing Cloudera Data Science Workbench (CDSW) hosts. If you need to repurpose any CDSW nodes for Cloudera AI, you must perform several tasks before starting the CDSW to Cloudera AI migration.

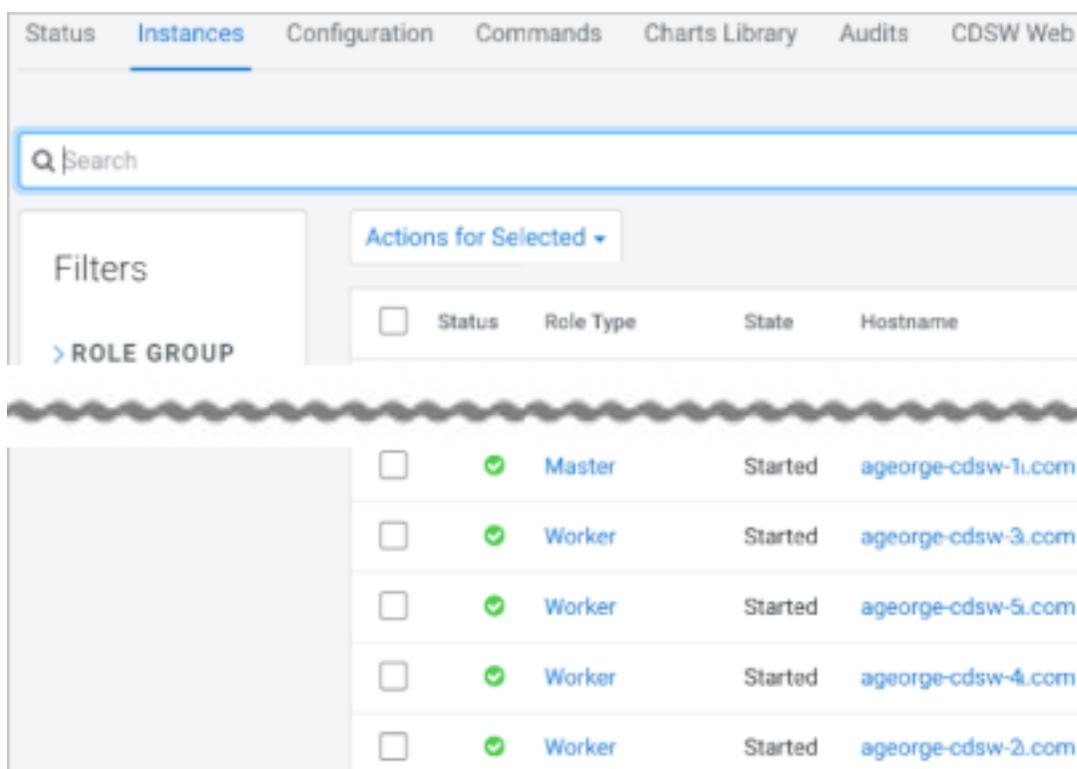
The pre-migration example is based on five hosts. In Cloudera Manager, the hosts appear in the CDSW service Status Summary:



The screenshot shows the 'Status' tab of the Cloudera Manager interface. It displays a 'Health Tests' section with a 'Show 5 Good' button. Below that is a 'Status Summary' table with the following data:

Component	Status
Application	1 Good Health
Docker Daemon	5 Good Health
Master	1 Good Health
Worker	4 Good Health
Hosts	5 Good Health

In Instances, in Role Type, you see the following role types.



The image presents how to repurpose two worker nodes named ageorge-cds-4.com and ageorge-cds-5.com:

1. Remove cluster hosts and Kubernetes nodes.
2. Install and configure the Cloudera on premises cluster on the repurposed nodes.

## Removing cluster hosts and Kubernetes nodes

Follow the instructions for removing cluster hosts and Kubernetes nodes.

### Before you begin

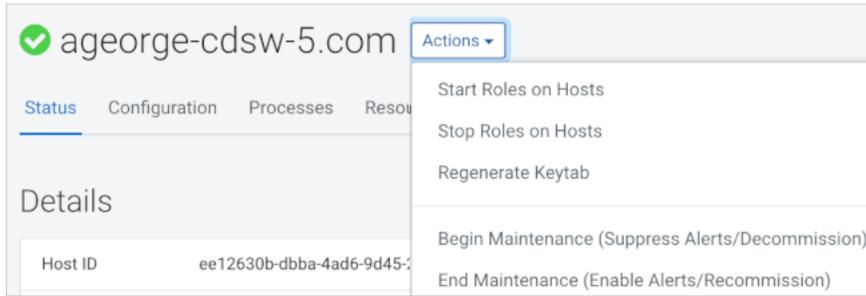
Before removing the node, shut down the nodes following the instructions in [Performing maintenance of a single host in the Embedded Container Service cluster](#).

### About this task

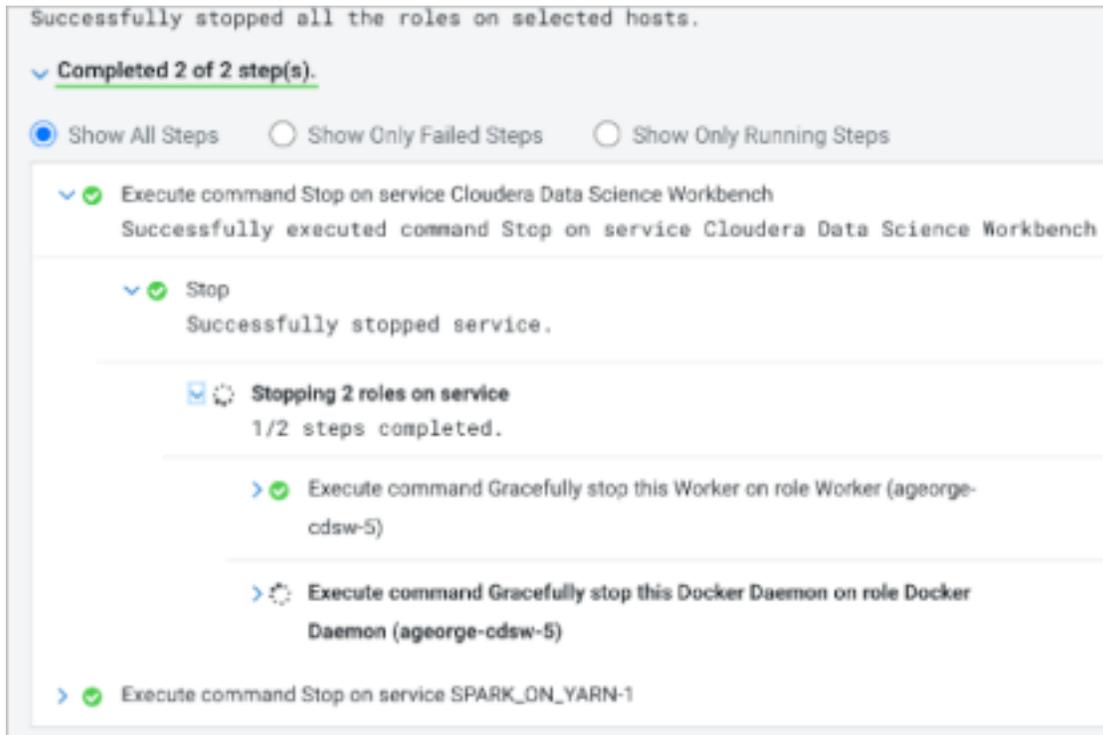
If you want to repurpose nodes from Cloudera on premises Base, you must use a different procedure from the one described below. For more information, see [Deleting hosts Cloudera on premises](#).

## Procedure

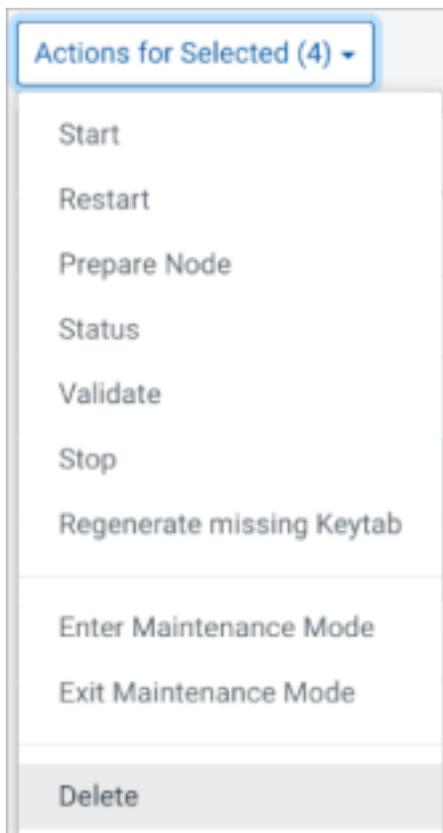
1. In Cloudera Manager, go to CDSW service Status , select the name of a host, and click Actions Stop Roles on Hosts .



After you click Confirm, you see a success message.



- In Instances, select the Docker Daemon and Worker roles assigned to ageorge-cds-4.com and ageorge-cds-5.com, and click Actions Delete .



In Instances, in Hostnames, ageorge-cds-4.com and ageorge-cds-5.com are no longer listed.

- Use kubectl commands to delete the nodes from Kubernetes, and then check that the nodes are removed.

```
[root@ageorge-cds-1 ~]# kubectl delete node ageorge-cds-4.com
node "ageorge-cds-4.com" deleted
[root@ageorge-cds-1 ~]# kubectl delete node ageorge-cds-5.com
node "ageorge-cds-5.com" deleted
```

```
[root@ageorge-cds-1 ~]# kubectl get nodes
NAME                STATUS    ROLES    AGE   VERSION
ageorge-cds-1.com   Ready    master   61m   v1.19.15
ageorge-cds-2.com   Ready    <none>   61m   v1.19.15
ageorge-cds-3.com   Ready    <none>   61m   v1.19.15
```

- Log into each host, ageorge-cds-4.com and ageorge-cds-5.com, and reset Kubernetes.

```
kubeadm reset
```

Expect a similar output:

```
[reset] WARNING: Changes made to this host by 'kubeadm init' or 'kubeadm
join' will be reverted.
[reset] Are you sure you want to proceed?
```

- Respond yes to the prompt, and follow the instructions in the resulting output.

Expect a similar output:

```
[preflight] Running pre-flight checks
...
```

The reset process does not clean CNI configuration. To do so, you must remove `/etc/cni/net.d`

The reset process does not reset or clean up iptables rules or IPVS tables.

If you wish to reset iptables, you must do so manually by using the "iptables" command.

If your cluster was setup to utilize IPVS, run `ipvsadm --clear` (or similar) to reset your system's IPVS tables.

The reset process does not clean your kubeconfig files and you must remove them manually.

Please, check the contents of the `$HOME/.kube/config` file.

**6.** Check that the Kubernetes nodes are removed.

```
kubectl get nodes
```

Expect a similar output:

NAME	STATUS	ROLES	AGE	VERSION
ageorge-cdsw-1.com	Ready	master	65m	v1.19.15
ageorge-cdsw-2.com	Ready	<none>	64m	v1.19.15
ageorge-cdsw-3.com	Ready	<none>	64m	v1.19.15

7. If the Spark Gateway role is active, in the SPARK ON YARN service, in Instances, select each Gateway role, and click Actions Delete .  
In this example, the Role Types for ageorge-cds-5.com and ageorge-cds-4.com hosts are selected.

The screenshot shows the Cloudera AI interface for the SPARK\_ON\_YARN-1 service. The 'Instances' tab is active, and a filter for 'ROLE TYPE' is set to 'Gateway'. A dropdown menu for 'Actions for Selected (2)' is open, with 'Delete' highlighted. The table below shows the following instances:

Role Type	State	Hostname
Gateway	Stopped	ageorge-cds-1.com
Gateway	Stopped	ageorge-cds-3.com
Gateway	Stopped	ageorge-cds-5.com
Gateway	Stopped	ageorge-cds-4.com
Gateway	Stopped	ageorge-cds-2.com
Gateway	N/A	cml-pvc-ocd-1.cr.hwx.com

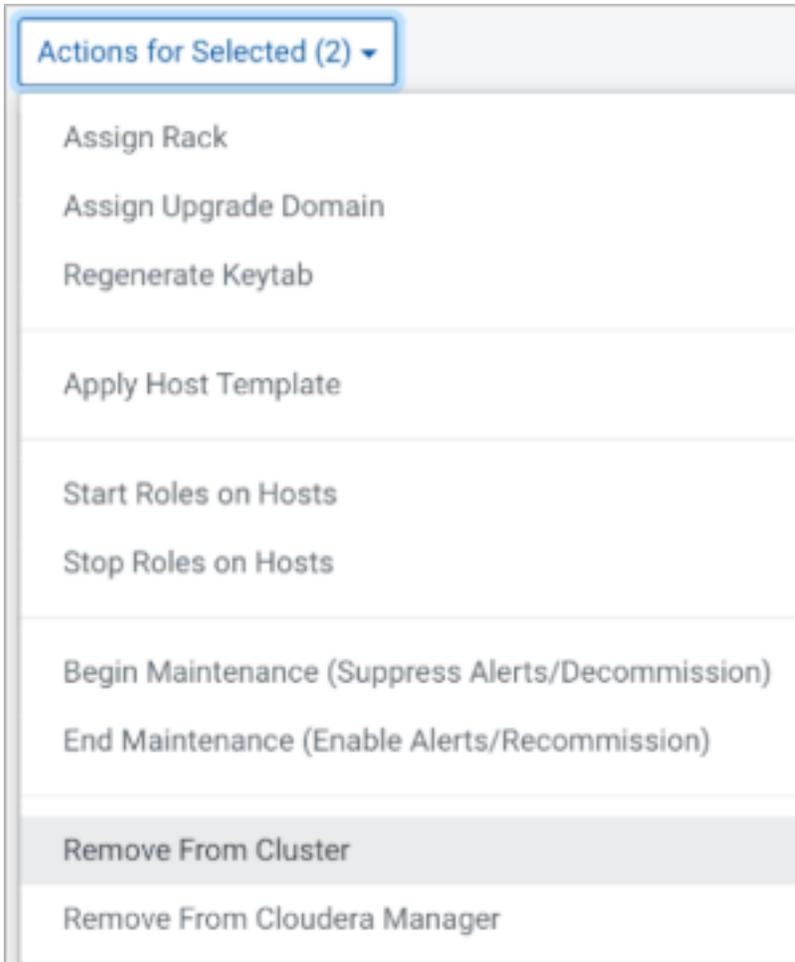
The screenshot shows the 'Delete Role Instances' dialog box. It asks 'Are you sure you want to delete the following role instances?' and shows a table with the following data:

Host	Role Instances
ageorge-cds-[4-5].com	DD W

The 'Delete' button is highlighted in red.

- Click Delete, and the Role Types assigned to ageorge-cds-5.com and ageorge-cds-4.com hosts are deleted.
8. Delete any other active roles from the hosts.

9. In All Hosts, select the ageorge-cds-5.com and ageorge-cds-4.com hosts, and click **Actions Remove from Cluster**.



You are prompted to confirm removing the hosts from the cluster:

Remove Hosts From Cluster

Removing these hosts will stop and delete all roles running on them and then remove them from their clusters. The hosts will still be managed by Role data directories will not be deleted.

Host
ageorge-cds-[4-5].com

Decommission Roles (**Warning: Removing the hosts without decommissioning the roles running on them can result in permanent data loss.**)

Skip Management and Authentication Service Roles

[Cancel](#) [Confirm](#)

Click **Confirm**, and view the success message upon completion.

### What to do next

Following this operation, add the node to the Cloudera Embedded Container Service cluster. See the details in [Installing Cloudera Private Cloud Data Services using Cloudera Embedded Container Service](#)

## Customizing CDSW for migrating host mounts

For security reasons, Cloudera AI does not allow you to mount a directory directly from hosts. You need to customize Cloudera Data Science Workbench (CDSW) runtime to make contents of a directory available to your Cloudera AI workloads.

### About this task

Before migration, you must perform a few pre-migration steps if you [mounted additional dependencies](#) from the host in CDSW. For example, you might have mounted directories containing libraries needed for running your workloads. You need to make these libraries available in Cloudera AI. In the pre-migration steps below, you set up CDSW for the migration to mount your libraries in Cloudera AI.

If you loaded libraries from a host mount in the past, Cloudera recommends you create a custom runtime in CDSW, change the project to use the new custom runtime, and then do the migration. However, for anything other than the libraries, load the data to all the sessions in Cloudera AI using the [custom runtime addons](#) procedure after migration to mount data in all the workloads in Cloudera AI. Custom runtime addons do not allow writes to the file system as the host mount in CDSW does.

### Procedure

1. [Create a customized ML Runtime](#).
2. If libraries were loaded from the host mount, configure your CDSW project to use the custom runtime by [adding the custom runtime](#) to CDSW before migration.



**Note:** After migration, use the custom runtime addons procedure mentioned above to mount anything other than libraries.

Libraries you add to the custom runtimes will be available to the Cloudera AI projects using that custom runtime.

## Using the CDSW to Cloudera AI Migration tool

A step-by-step procedure covers how to migrate Cloudera Data Science Workbench (CDSW) 1.10.0 and higher version on premises to Cloudera AI. You use the UI-driven migration tool to migrate your workload automatically from your deployed CDSW, which you installed on the same cluster as Cloudera AI.

### About this task

In this task, you automatically migrate CDSW 1.10.0 or later cluster to Cloudera AI in on premises 1.5.3 or later. You can expect some downtime, which is proportional to the volume the workloads you have to migrate.

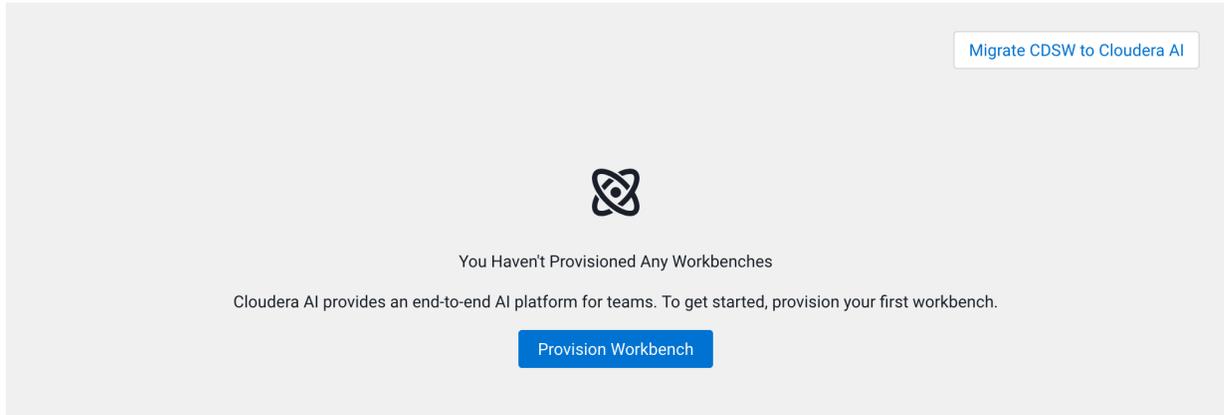
### Before you begin

Meet prerequisites as described in the previous topic before starting migration.

## Procedure

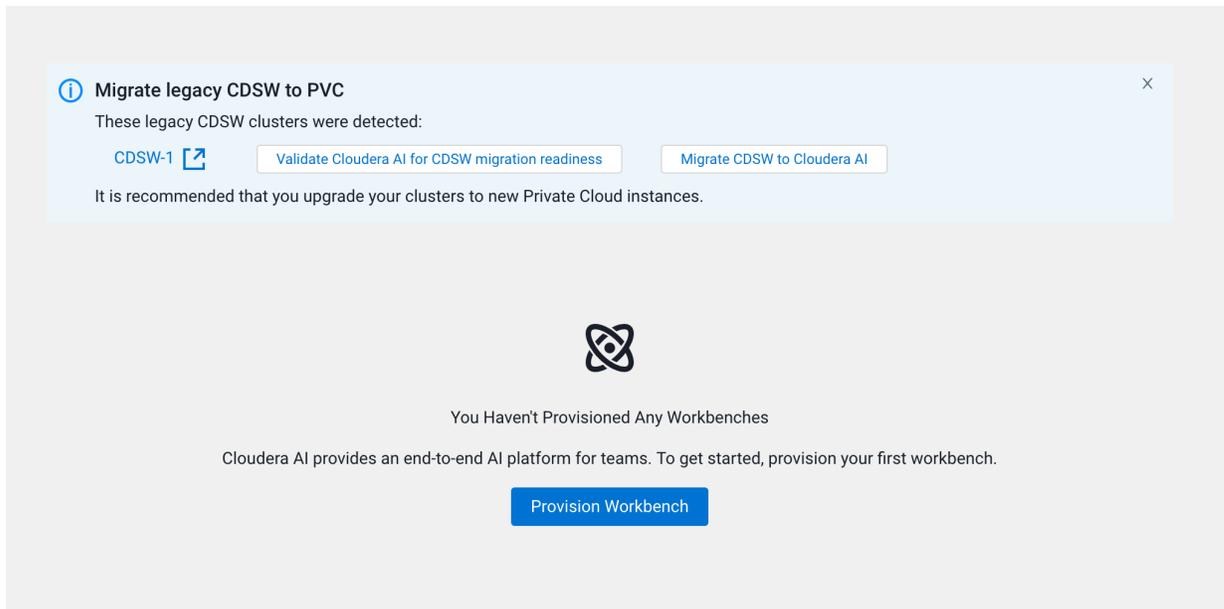
1. Log into Cloudera on premises, and navigate to Cloudera AI Workbenches .  
The system detects the presence of your legacy CDSW installation and provides a button to migrate CDSW to Cloudera AI.

Cloudera AI Workbenches



2. Click Migrate CDSW to Cloudera AI.

The Migration tool provides an option to check Cloudera AI for CDSW migration readiness or just continue with the migration.



3. Click the option for how you would like to proceed.
  - Check Cloudera AI for CDSW migration readiness - The readiness or preflight check creates a new workbench and runs readiness checks on the workbench prior to performing the migration.
  - Migrate CDSW to Cloudera AI - The Migration tool provides a Cloudera AI Workbench provision window with additional options.

4. If you choose the Check Cloudera AI for CDSW migration readiness option, the Migration tool displays the validation page.

## Provision Cloudera AI Workbench

Provision an on-demand Cloudera AI Workbench.

\* Workbench Name

\* Select Environment

Environment type: **None selected**

\* Namespace ⓘ

NFS Server ⓘ

Internal  External

ⓘ This selection uses an external NFS export path (or a subdirectory within it).

\* Existing NFS ⓘ

ⓘ Note: An administrator must run **chown 8536:8536** on the NFS directory.

ⓘ The directory must be empty and not used by another workbench.

NFS Protocol version ⓘ

### Production Machine Learning

Enable Governance ⓘ

Enable Model Metrics ⓘ

### Other Settings

Enable TLS ⓘ

Enable Monitoring ⓘ

Cloudera AI Static Subdomain ⓘ

- a) If you would like to provide a Kubeconfig for the migration check, click File Upload, then Choose File, and select the Kubeconfig file.

The Kubeconfig file can be found at `/etc/kubernetes/admin.conf` on the CDSW cluster.

If you cannot access `/etc/kubernetes/admin.conf` from the UI as instructed in the previous step, download the file from your CDSW cluster to your local machine, and then try to select the Kubeconfig file from the UI again.

- b) In the Migration timeout section, accept the default 24 hours timeout, or if your CDSW workload is hundreds of gigabytes, increase the migration time up to 336 hours (14 days).

Increasing the migration timeout value does not cause a delay in the migration of a small workload.

- c) In the Workbench Name field, type an arbitrary name.  
d) In the Select Environment field, select your Cloudera environment.

After the readiness check has completed, the Migration tool displays a status. You can see the readiness check summary in the Workbench Details page.

- If the readiness check fails, you can obtain additional information about the failure on the Workbench Details page.

Status	Version	Workbench	Environment	Creation Date	Cloud Provider	Actions
Migration Readiness Failed	2.0.38	preflight	eah792-env	04/13/2023 12:13 PM EDT	ECS	⋮

- After you've addressed the issue resulting in the failed readiness check, you can retry the readiness check by choosing Retry Migration Readiness check from the Actions menu.

Status	Version	Workbench	Environment	Creation Date	Cloud Provider	Actions
Migration Readiness Failed	2.0.38	preflight	eah792-env	04/10/2023 12:13 PM EDT	ECS	⋮

- View Workbench Details
- View Event Logs
- Manage Access
- Open GrafanaUI
- Upgrade Workbench
- Backup Workbench
- Remove Workbench
- Retry CDSW migration
- Incremental CDSW migration
- Retry Migration Readiness Check

- e) After the readiness check, incremental migration can be performed to continue the migration operation. Alternatively, you can choose to create a new workbench by clicking the Migrate CDSW to Cloudera AI button.

5. When you proceed with the CDSW to Cloudera AI migration, the Migration tool displays the Migrations Settings window.

## Provision Cloudera AI Workbench

Provision an on-demand Cloudera AI Workbench.

\* Workbench Name

\* Select Environment

Environment type: **None selected**

\* Namespace ⓘ

NFS Server ⓘ

Internal  External

ⓘ This selection uses an external NFS export path (or a subdirectory within it).

\* Existing NFS ⓘ

ⓘ Note: An administrator must run **chown 8536:8536** on the NFS directory.

ⓘ The directory must be empty and not used by another workbench.

NFS Protocol version ⓘ

### Production Machine Learning

Enable Governance ⓘ

Enable Model Metrics ⓘ

### Other Settings

Enable TLS ⓘ

Enable Monitoring ⓘ

Cloudera AI Static Subdomain ⓘ

- If you would like to provide a Kubeconfig for the migration, click File Upload, then Choose File, and select the Kubeconfig file.

The Kubeconfig file can be found at `/etc/kubernetes/admin.conf` on the CDSW cluster.

If you cannot access `/etc/kubernetes/admin.conf` from the UI as instructed in the previous step, download the file from your CDSW cluster to your local machine, and then try to select the Kubeconfig file from the UI again.

- In the Migration timeout section, accept the default 24 hours timeout, or if your CDSW workload is hundreds of gigabytes, increase the migration time up to 48 hours.

Increasing the migration timeout value does not cause a delay in the migration of a small workload.

- In the Workbench Name field, type an arbitrary name.

- In the Select Environment field, select your Cloudera environment.

- Accept default values for other options, and click Provision Workbench.

After the Cloudera AI installation, the migration readiness checks and the migration follow automatically. Status indicators show the progress of the installation and migration. During the migration, you can access the CDSW cluster. The migration process does not stop CDSW pods. The Cloudera AI Workbench is stopped.



**Note:** Any changes to CDSW during the migration will not be copied to Cloudera AI. These changes will be copied in subsequent incremental migrations.

- To display the progress of the migration including events and logs while the workbench is in migration mode, navigate to the Workbench Details page and click the Migration Progress tab.

You can also view details of the migration and events and logs by clicking the appropriate tabs.

Cloudera AI Workbenches / migrate

Status  Migration Started

Details Events & Logs Migration Progress Actions

Migration Progress

CDSW to Cloudera AI migration Migration Start Time: 10/4/2023 16:26:14

Migrating CDSW cluster to Cloudera AI

- Running preliminary sanity checks
- Verifying kubemetes configuration file
- Running preflight check
- Checking the migrations
- Checking the auth type
- Getting docker credentials
- Scale down CDSW deployments
- Wait for database affecting job
- Scaledown Cloudera AI deployments

- When the initial migration is complete, the state changes to Validate Migration Started.

One or more workbenches are in 'Validate Migration Started' state. Any changes done to the workbench will be overwritten during the incremental migration.

Search Workbenches Environment All Provision Workbench

Status	Version	Workbench	Environment	Creation Date	Cloud Provider	Actions
Validate Migration Started	2.0.39	aq8z4migration1	aq8z4-env	04/26/2023 6:25 PM EDT	ECS	

Displaying 1 - 1 of 1 < 1 > 25 / page

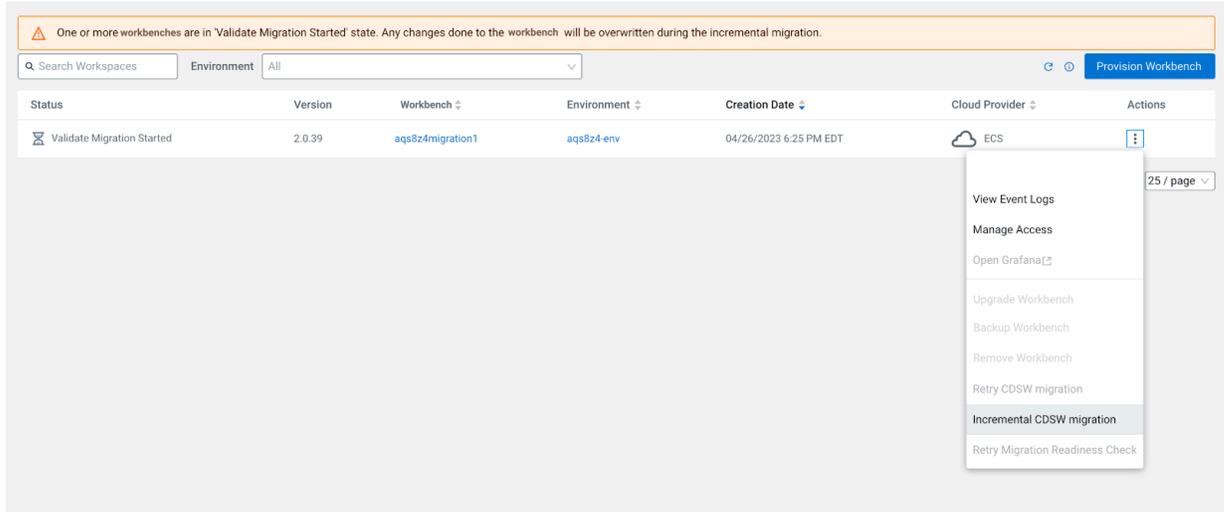
- Open the workbench by clicking the workbench name and validate the workloads.

Any changes made to the workbench while you are validating the workbench will be overwritten during the incremental migration.

14. At this point, you can choose to do multiple incremental migrations or a single, longer migration.

- To perform incremental migrations, select the Incremental CDSW Migration option from the Actions menu. CDSW is not stopped during incremental migrations.
- To perform a single, longer migration, select the Incremental CDSW Migration option from the Actions menu, click the Perform final migration checkbox, then click OK.

During the final migration, CDSW will be stopped and will not be restarted. After the final migration, only the Cloudera AI Workbench will be active.



The Migration tool displays the Incremental CDSW migration dialog box.

## Incremental CDSW migration X

---

Incremental CDSW migration to workbench : m4396nmigrate

Kubeconfig ⓘ

File Upload  Direct Input

Choose File

Migration timeout (in hours) ⓘ

02448

02448

Perform final migration

---

Cancel OK

15. If you choose to perform incremental migrations, choose your parameters for the migrations.

**CDSW Kubeconfig**

This is an optional parameter. If there is no change to the CDSW Kubeconfig, then you do not need to select this option. The system will use the kubeconfig that was provided during the initial migration. This option can be helpful if CDSW is restarted during the migration and there are changes to the CDSW kubeconfig.

**Migration timeout**

You can specify the amount of time allowed for the migration before it timeouts. This timeout value is specified in hours and can range from 0 to 48 hours.

16. If you choose to perform incremental migrations, when the Cloudera AI validation is complete and all workloads work as expected, select Perform final migration and choose the appropriate option.

## Incremental CDSW migration ✕

---

Incremental CDSW migration to workbench : m4396nmigrate

Kubeconfig  ⓘ

File Upload
  Direct Input

Choose File

Migration timeout (in hours)  ⓘ

0 24 48



Perform final migration

- Stop applications
- Stop jobs
- Stop models

⚠ Do final migration of CDSW to this Cloudera AI workbench. In this case, source CDSW workbench will be tuned off after migration and workloads on migrated Cloudera AI workbench will start running.

Cancel

OK

After you perform the final migration, CDSW will be in a stopped state.

Perform final migration provides three options:

#### Stop applications

If you select the Stop applications option, the applications in the Cloudera AI will be in the stopped state after the migration. You must manually start each application after the final migration.

#### Stop jobs

If you select the Stop jobs option, the recurring jobs in the Cloudera AI will be in the paused state after the migration. You must manually start each job after the final migration.

### Stop models

If you select the Stop models option, the models in the Cloudera AI will be in the stopped state after the migration. You must manually start each model after the final migration.

17. Now that the migration is complete, you can use Cloudera AI.

## Troubleshooting preflight migration check issues

Consider the list of preflight check issues and workarounds.

### Config Readiness

In Cloudera Data Science Workbench (CDSW), we are using KubeDNS for handling the DNS. In Cloudera AI we are using core DNS. If you add any custom KubeDNS configurations in the CDSW, those will not be copied to Cloudera AI automatically. This preflight check verifies if there is any custom configuration added in CDSW. This preflight check will flag the same if the custom configurations are detected.

Troubleshooting failures:

Note any custom configurations added to the KubeDNS. Check if these are really necessary for Cloudera AI.

### Registry Readiness

This preflight check will verify the registry readiness in Cloudera AI. It will check if the created Cloudera AI cluster has proper registry permissions.

Troubleshooting failures:

Check if the secret `cdp-private-installer-embedded-registry-docker-pull-secret` in the control plane namespace is present in Cloudera AI. If the configuration is present, check if the docker configuration in this secret is correct.

### Host Mount Readiness

CDSW supports host mount. But it is not supported in Cloudera AI. It is not possible to migrate the workloads containing the host mounts.

Troubleshooting failures:

The workloads containing the host mounts should be modified in CDSW before the migration. The users may convert the workloads using the engines to runtimes.

### Engine Type Readiness

Engine is deprecated in Cloudera AI. If the CDSW contains any workloads using the engine or custom engines, those workloads must be converted to runtime before the migration.

Troubleshooting failures:

The workloads based on engine or custom engine should be converted to runtime-based workloads.

### NFS Filesystem Readiness

The CDSW to CML migration tool supports migrating project files from the CDSW internal NFS server to the CML project storage. This readiness check will verify the NFS filesystem size for the migration.

Troubleshooting failures:

Please check the storage configuration of the NFS storage in Cloudera AI. Please ensure enough storage is in the NFS storage to do the migration.

## Runtime Addons

Cloudera AI supports many runtime add-ons such as Spark, Hadoop, CLI, etc. This preflight check verifies that all the runtime add-ons are installed in the Cloudera AI properly.

Troubleshooting failures:

In Cloudera AI, go to [Site Administrator Runtime](#) and check the status of the runtime addons. If any of the runtime addons are in the wrong state, click actions and reload it.

Site Administration / Runtime Q Projec

Overview Users Teams Usage Quotas Models **Runtime** Security AMPs Learning Hub Settings

Hadoop CLI Version

**Runtime Addons**

Status	Name	ID	Component	Created At	Reason	Actions
Available	Hadoop CLI - CDP 7.1.7-2000	2	HadoopCLI	05/03/2023 1:59 PM		⋮
Available	Hadoop CLI - CDP 7.1.8	1	HadoopCLI	05/03/2023 1:59 PM		⋮
Available	Ozone - 718.2.0-b38	3	Ozone	05/03/2023 1:59 PM		⋮
Available	Spark 2.4.7 - CDP 7.1.7.2000	4	Spark	05/03/2023 1:59 PM		⋮
Available	Spark 3.2.3 - CDP 7.1.7.2000	5	Spark	05/03/2023 1:59 PM		⋮

< 1 >

## Service Readiness

This preflight check verifies all the services active in CDSW in enabled and started in Cloudera AI.

Troubleshooting failures:

Ensure the necessary services are started in Cloudera AI during the workspace provisioning.

## Versions Readiness

The migration of CDSW to Cloudera AI is supported only from CDSW version 1.10.0. This preflight check will verify the source CDSW version.

Troubleshooting failures:

Please update the CDSW to version  $\geq 1.10.0$ .

## Limitations of migrating CDSW to Cloudera AI

You need to avoid unsupported operations that might cause migration to fail.

The following operations are not supported for migration of Cloudera Data Science Workbench (CDSW) 1.10.0 or higher version to Cloudera AI:

- The migration of CDSW to Cloudera AI is not supported for the Cloudera Embedded Container Service cluster installed with the internal registry alias option.
- Migration to Cloudera on premises OpenShift Container Platform (OCP) is not supported.
- Custom configurations, such as host or Kubernetes configurations, are not migrated. You must note these configurations and manually configure your on premises cluster after migration.
- CDSW and Cloudera on premises clusters must run side-by-side during migration.

- CDSW projects that use engines with Spark might not work as expected after migration.
- Migrating the sessions created with a custom engine in CDSW will not work in the migrated Cloudera AI as the engine architecture differs between Cloudera AI and CDSW. You must move from the custom engine to a custom ML Runtime before the migration.
- CDSW projects that access HBase might not work after migration.

### Related Information

[Creating customized ML Runtimes](#)

## Post-migration tasks

After migrating Cloudera Data Science Workbench (CDSW) 1.10.0 or higher version to Cloudera AI, you need to perform several tasks before moving users to Cloudera AI.

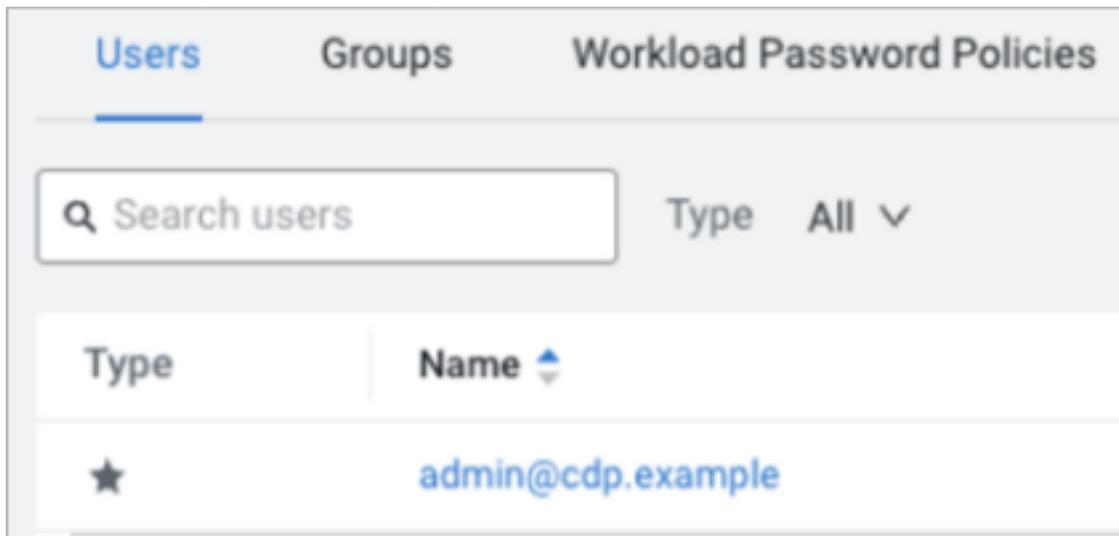
### About this task

You begin this task by assigning the user, group and the roles in the on premises control plane.

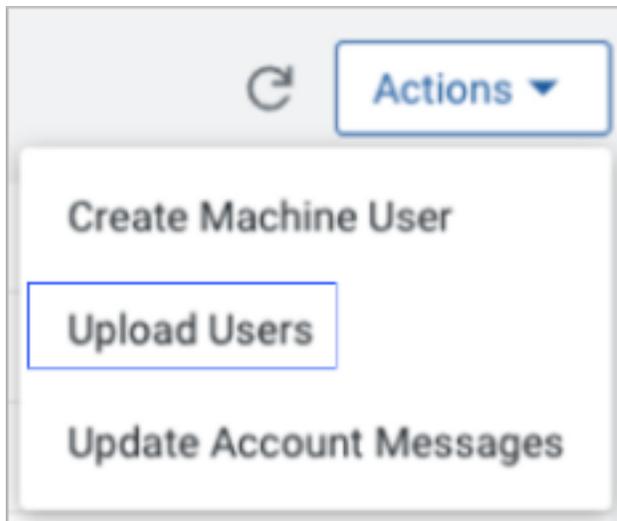
The CDSW to Cloudera AI migration in Cloudera Manager updates the docker registry, engines, and model builds. You follow steps to assign roles to users and groups, import Grafana dashboards you previously exported, configure endpoints and DNS resolution, and configure LDAP.

### Procedure

1. In Cloudera Manager, click **User Management** , and select **Users**.



2. Click **Actions Upload Users** .



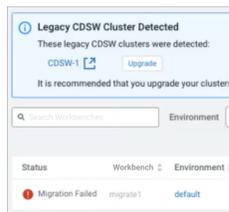
3. Select Groups, click Create Groups, and create groups.
4. Start the Cloudera AI workspace on the Cloudera AI cluster, and check workloads. Start new sessions, jobs, models, and applications. Try starting existing workloads that were migrated from CDSW.
5. Import the Grafana dashboards you exported earlier.
6. Configure cluster endpoint connectivity per the information about required connections you had in your CDSW cluster.
7. If you customized your DNS configuration on CDSW, manually configure your DNS in your on premises cluster. If you did not customize your DNS configuration, the migration tool sets up the default DNS configuration in your on premises cluster.
8. Configure [LDAP on Cloudera AI](#) [LDAP on Cloudera AI](#), and [grant user access on Cloudera AI](#). Cloudera AI in on premises supports only LDAP.
9. Disable the CDSW cluster, and give users access to Cloudera AI.

## Troubleshooting CDSW migration to Cloudera AI

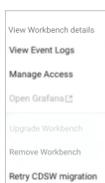
If your migration from Cloudera Data Science Workbench (CDSW) to Cloudera AI fails, you can restart migration. Also, learn how to find logs for debugging migration and other issues.

### Failed Migration

Problem: In the event your migration fails, the Migration failed indicator appears.



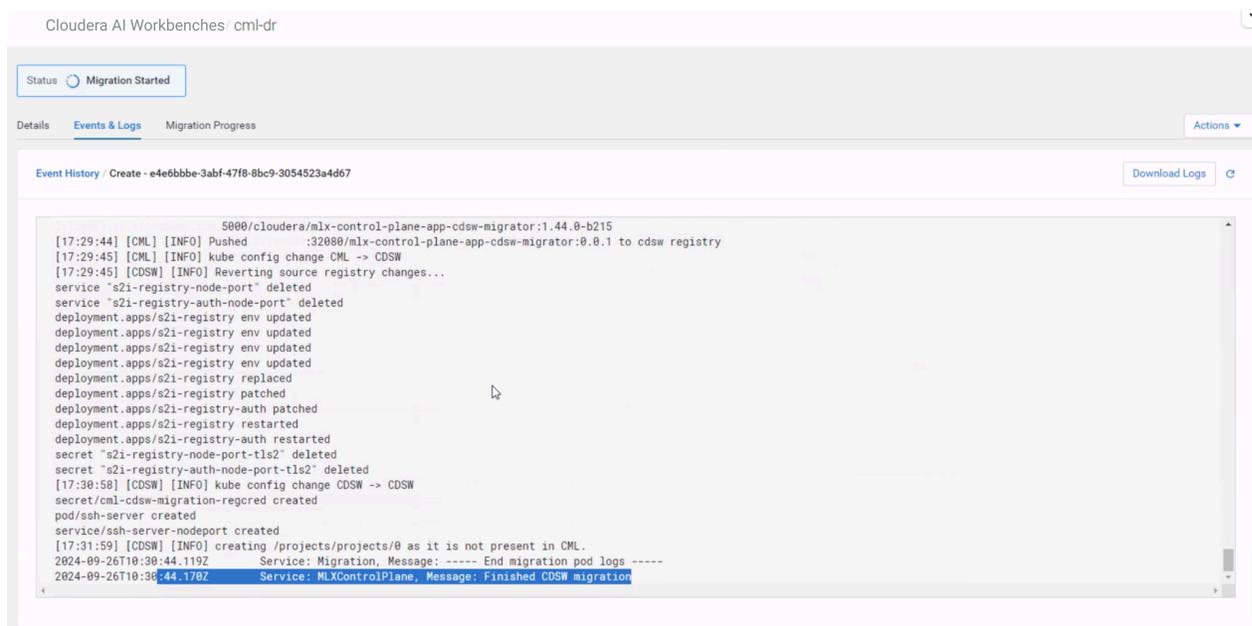
Solution: In workbenches, click **Options**  , and select Retry CDSW Migration:



## CDSW to Cloudera AI migration frozen with large Projects

**Problem:** If the Project being migrated from CDSW to Cloudera AI is large, it can get stuck, frozen, while waiting for all pieces of Project information to get migrated. After a while, the logs display that the migration was successful, however the UI is still unavailable:

### Figure 1: Successful migration with UI still unavailable



**Solution:** Instruct Cloudera manually that the migration has finished:

1. Find the Customer Resource Number (CRN) of the new workbench from the Control Plane.
2. Open a shell into the `cdp-embedded-db-0` pod which is in the Cloudera Data Platform namespace. This can be done from the Kubernetes dashboard or with `kubectl`.
3. Add a new entry into the database that tells the system that the migration is completed:

```
psql
\c db-mlx
INSERT INTO event(instance_id, resource_type, status, operation, user_id)
  SELECT id, 'mlx_instance', 'started', 'validateMigration', creatorcrn f
  rom mlx_instance where mlx_instance.crn='##<crn>###';
```

Following these steps, the Cloudera UI is updated, and displays the information that the migration must be validated. You can continue the migration process.

## Viewing logs

**Problem:** In the event your migration fails, and the retry also fails, you need to get information about the failure.

**Solution:** In workbenches, click Options , and select View Event Logs.

### Problem when migrating parcels to a new Cloudera AI Workbench

If parcels are located in a custom directory specified by `parcel_repo_home` in Cloudera Manager, instead of the default location (`/opt/cloudera/parcels`), then the CDSW to Cloudera AI migration script will not find them.

Workaround:

1. Check that the Cloudera Embedded Container Service parcels are located at `/opt/cloudera/parcels` and not at a custom location.
2. If not, then copy the existing parcels to the default location:

```
cp -rp <custom-ecs-parcel-path>/ECS* /opt/cloudera/parcels/
```