

CDP.1.5.0.

CDSW to CML migration

Date published: 2022-12-14

Date modified:

CLOUDERA

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

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

Migrating Data Science Workbench (CDSW) to Machine Learning.....	4
Prerequisites for CDSW to CML migration.....	4
Repurposing CDSW nodes for CML.....	6
Cordoning and draining Kubernetes nodes.....	7
Removing cluster hosts and Kubernetes nodes.....	8
Installing CML.....	12
Customizing CDSW for migrating host mounts.....	15
Using the CDSW to CML Migration tool.....	15
Troubleshooting Preflight Migration Check Issues.....	30
Limitations of CDSW to CML migration.....	33
Post-migration tasks.....	33
Troubleshooting CDSW to CML migration.....	35

Migrating Data Science Workbench (CDSW) to Machine Learning

If you use Cloudera Data Science Workbench (CDSW) 1.10.0 and later on premises, consider migrating to Cloudera Machine Learning (CML) 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 CML Private Cloud. Next, you install CML on the hosts. Finally, the UI-driven migration tool runs scripts in the background to migrate your workload automatically after installing CML on the cluster with your deployed CDSW.

Prerequisites for CDSW to CML migration

Before migrating from Cloudera Data Science Workbench (CDSW) to Cloudera Machine Learning (CML) in Cloudera Data Platform Private Cloud, you must meet a number of prerequisites to succeed. A prerequisite for migration is the installation of Cloudera Machine Learning 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 Machine Learning:

Table 1: Supported migration versions for Cloudera Data Science Workbench and Cloudera Machine Learning

Supported CDSW versions	Target CML PVC version
CDSW 1.10.0 CDSW 1.10.1 CDSW 1.10.2	1.5.1
CDSW 1.10.5	1.5.2 1.5.3 1.5.4



Note:

With CDSW 1.10.3 or 1.10.4, an upgrade to CDSW 1.10.5 is recommended and a migration to Cloudera Machine Learning Private Cloud to a version higher than 1.5.2.

Migration from CDSW, configured with LDAP, SAML, or LOCAL authentication to Cloudera Machine Learning, 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 Machine Learning, as well as downloading CDSW-related Grafana dashboards, so you can upload them after migration to Cloudera Machine Learning.

Procedure

1. You must have a CDSW 1.10.0 or later version cluster in Cloudera Data Platform; otherwise, choose one of the following options:

- If you have a CDSW installation in either CDH or HDP, migrate to Private Cloud 1.5.1 or later version, and then migrate CDSW to Cloudera Machine Learning.
- If you have CDSW installation earlier than 1.10.0, upgrade to CDSW 1.10.0 or later versions.

2. If you do not have LDAP set up in your CDSW cluster on Cloudera Data Platform, 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 Machine Learning software requirements for Private Cloud, including storage, for installing Cloudera Machine Learning on Cloudera Data Platform Private Cloud 1.5.1 or later version. For Cloudera Machine Learning software requirements for Private Cloud, see [CML software requirements for Private Cloud](#).
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, note your custom settings.

You need to use this information after migration to set up endpoints in your Private Cloud 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 Private Cloud cluster.

8. Gather information about your LDAP configurations on CDSW.

After migration, you must set up LDAP again on the Cloudera Machine Learning 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 Machine Learning.

Cloudera Machine Learning uses the core-DNS, which is incompatible with the CDSW Kube-DNS.

10. In Cloudera Manager, select install and upgrade to CDP Private Cloud 1.5.1 or later version using the Embedded Container Service on your CDSW cluster.

Migration of your CDSW workloads to Cloudera Machine Learning on OpenShift is not supported.

11. During the installation of Cloudera Data Platform Private Cloud Data Services using Embedded Container Service set up a network connection between CDSW and the Cloudera Data Platform Private Cloud cluster if you select Airgap.

12. Enable those Cloudera Machine Learning features during installation that you were using in CDSW.

For example, enable model metrics and monitoring.

Production Machine Learning

☐ Enable Governance ⓘ

☒ Enable Model Metrics ⓘ

Other Settings

☐ Enable TLS ⓘ

☒ Enable Monitoring ⓘ

If you do not enable the same, or similar, Cloudera Machine Learning features during installation that you were using in CDSW, you will not be able to use the Cloudera Machine Learning 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](#)

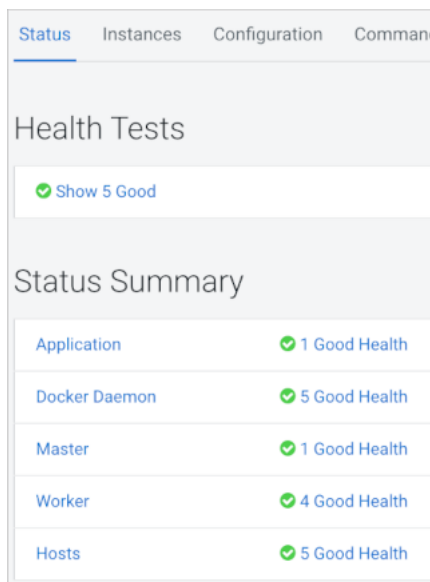
[CML software requirements for Private Cloud](#)

[CML software requirements for Private Cloud](#)

Repurposing CDSW nodes for CML

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 Machine Learning (CML), you must perform several tasks before starting the CDSW to CML migration.

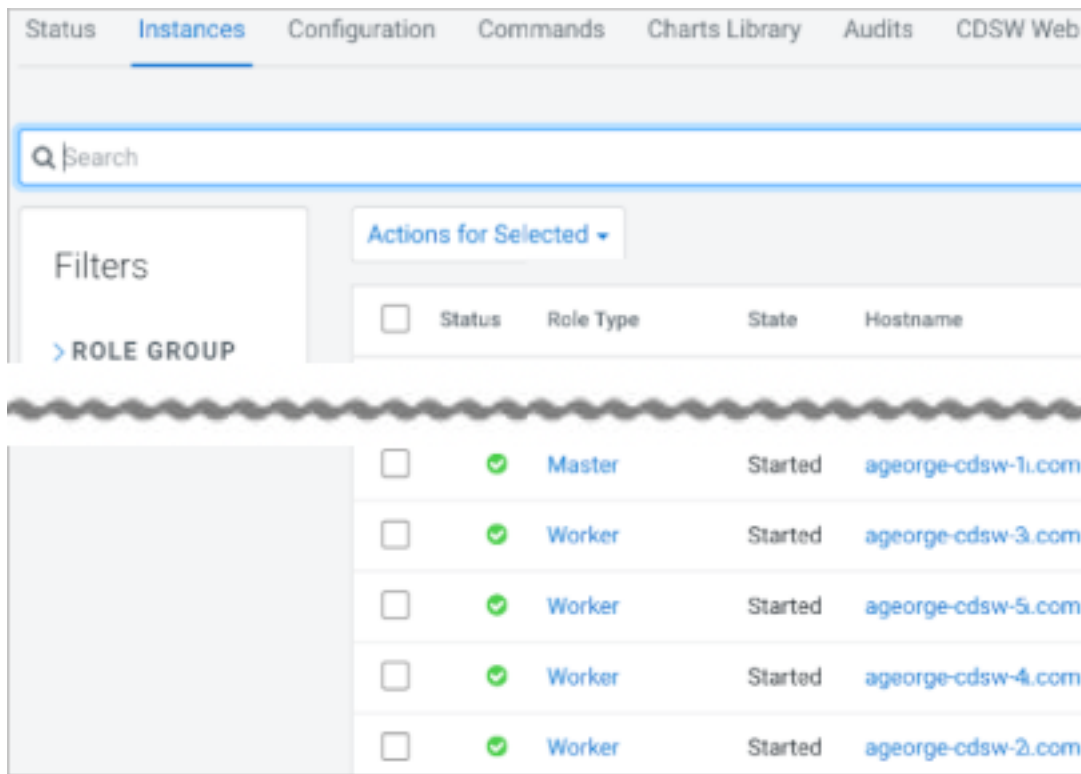
The pre-migration example is based on five hosts. In Cloudera Manager, the hosts appear in the CDSW service Status Summary:



The screenshot shows the 'Status' tab of the Cloudera Manager interface for the CDSW service. It displays a 'Health Tests' section with a 'Show 5 Good' button. Below this 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.



You see how to repurpose two worker nodes named ageorge-cds-4.com and ageorge-cds-5.com. First, you cordon and drain the Kubernetes nodes. Next, you remove cluster hosts and Kubernetes nodes. Finally, you install and configure the Cloudera Private Cloud cluster on the repurposed nodes.

Cordoning and draining Kubernetes nodes

You learn kubectl commands to cordon and drain the nodes in the Kubernetes cluster.

Procedure

Cordon and drain the nodes ageorge-cds-4.com and ageorge-cds-5.com from the Kubernetes cluster using kubectl commands.

```
[root@ageorge-cds-1 ~]# kubectl get nodes
NAME                                STATUS    ROLES    AGE   VERSION
ageorge-cds-1.com                   Ready     master   42m   v1.19.15
ageorge-cds-2.com                   Ready     <none>    42m   v1.19.15
ageorge-cds-3.com                   Ready     <none>    42m   v1.19.15
ageorge-cds-4.com                   Ready     <none>    42m   v1.19.15
ageorge-cds-5.com                   Ready     <none>    41m   v1.19.15

[root@ageorge-cds-1 ~]# kubectl cordon ageorge-cds-4.com
node/ageorge-cds-4.cc-os.cloudera.com cordoned

[root@ageorge-cds-1 ~]# kubectl cordon ageorge-cds-5.com
node/ageorge-cds-5.com cordoned

[root@ageorge-cds-1 ~]# kubectl get nodes
NAME                                STATUS              ROLES    AGE   VERSION
ageorge-cds-1.com                   Ready               master   47m   v1.19.15
ageorge-cds-2.com                   Ready               <none>    47m   v1.19.15
ageorge-cds-3.com                   Ready               <none>    47m   v1.19.15
ageorge-cds-4.com                   Ready,SchedulingDisabled <none>    47m   v1.19.15
```

```
ageorge-cds-5.com    Ready,SchedulingDisabled    <none>    46m    v1.19.15

kubectl drain ageorge-cds-4.com --ignore-daemonsets --delete-local-data --force

kubectl drain ageorge-cds-5.com --ignore-daemonsets --delete-local-data --force
```

Removing cluster hosts and Kubernetes nodes

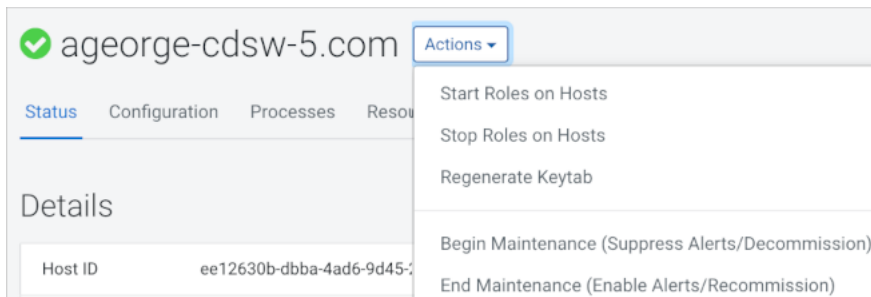
You stop roles assigned to the hosts you want to repurpose for the installation of Cloudera Private Cloud. Next, you delete nodes from Kubernetes. Finally, you go back to Cloudera Manager and remove hosts from the CDSW cluster.

About this task

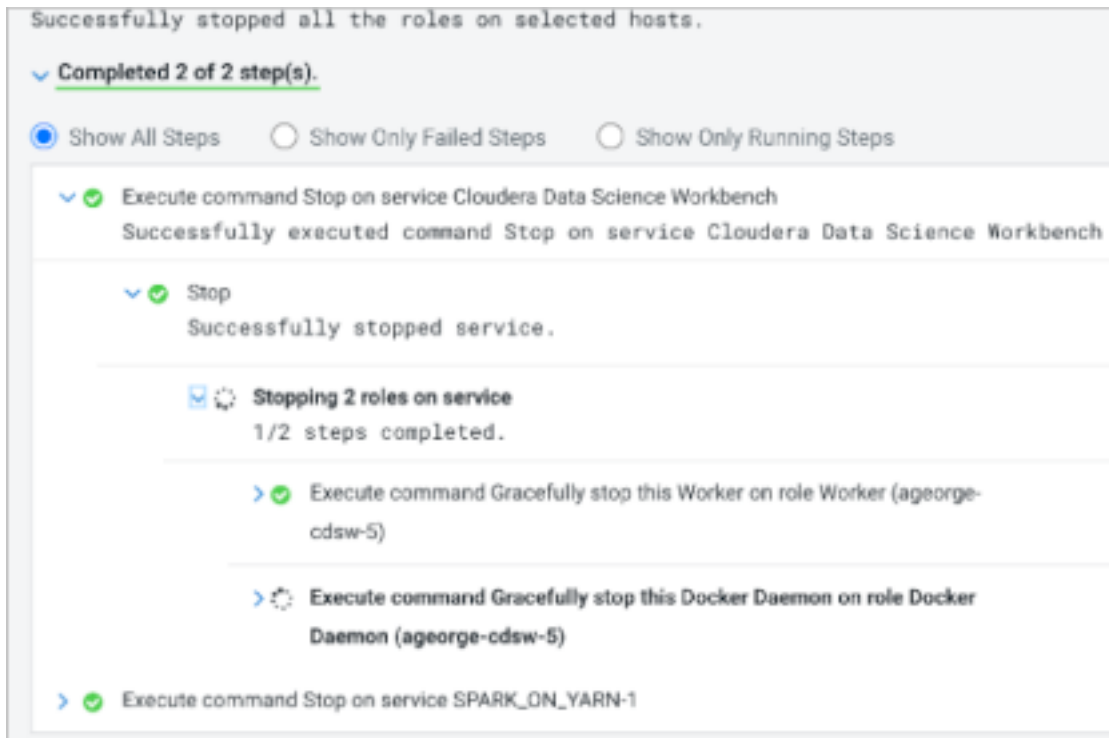
If you want to repurpose nodes from CDP Private Cloud Base, you must use a different procedure from the one described below. For more information, see [Deleting hosts \(CDP Private Cloud\)](#).

Procedure

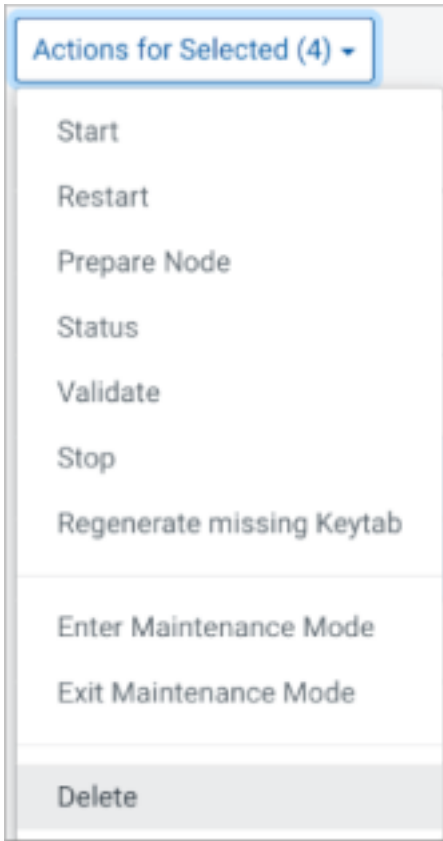
1. In Cloudera Manager, go to CDSW service Status , select the name of a host, and click Actions Stop Roles on Hosts .



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.

- Using kubectl commands, delete the nodes from Kubernetes, and then check that the nodes are gone.

```
[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
```

Output looks something like this:

```
[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 instructions in the resulting output.

Output looks something like this:

```
[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 gone.

```
kubectl get nodes
```

The output looks something like this:

NAME	STATUS	ROLES	AGE	VERSION
ageorge-cds-1.com	Ready	master	65m	v1.19.15
ageorge-cds-2.com	Ready	<none>	64m	v1.19.15
ageorge-cds-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-cdsw-5.com and ageorge-cdsw-4.com hosts are selected.

The screenshot shows the 'SPARK_ON_YARN-1' service page with the 'Instances' tab selected. On the left, filters are applied for 'STATUS' (Stopped) and 'ROLE TYPE' (Gateway). A table lists several Gateway roles, all in a 'Stopped' state. Two roles, 'ageorge-cdsw-5.com' and 'ageorge-cdsw-4.com', are highlighted. An 'Actions for Selected (2)' dropdown menu is open, showing options like Start, Restart, Stop, Regenerate missing Keytab, Enter Maintenance Mode, Exit Maintenance Mode, and Delete. The 'Delete' option is selected. Below the table, a 'Delete Role Instances' dialog box is displayed, asking for confirmation to delete the selected role instances. The dialog shows the host 'ageorge-cdsw-[4-5].com' and provides 'Cancel' and 'Delete' buttons.

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

Delete Role Instances

Are you sure you want to delete the following role instances?

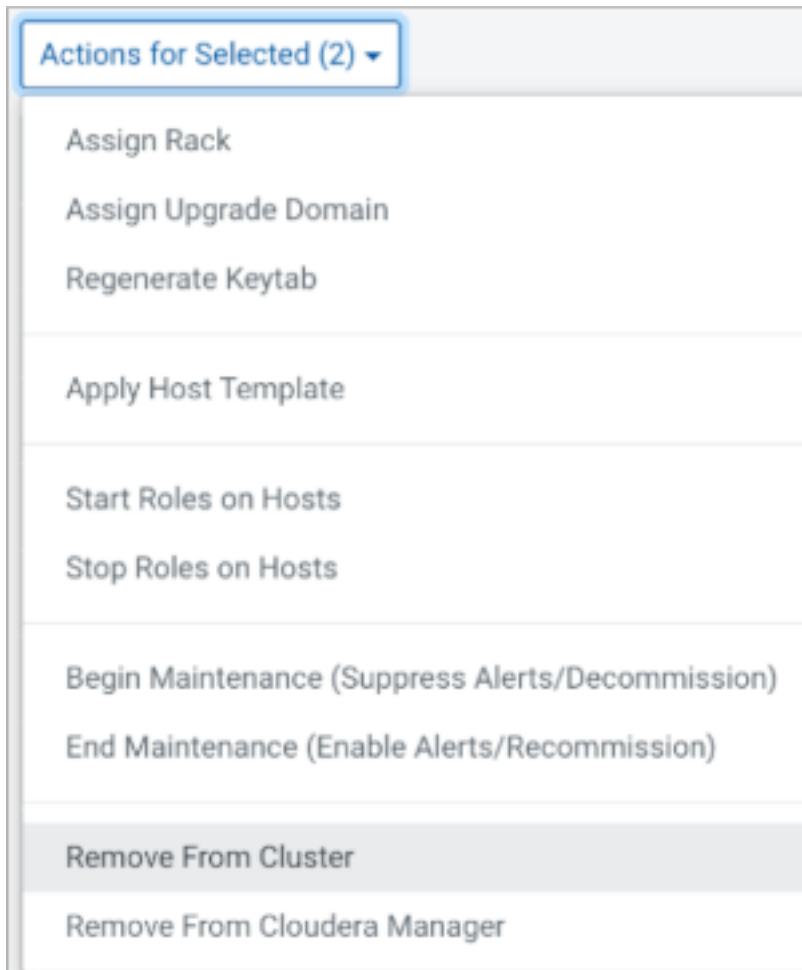
Host	Role Instances
ageorge-cdsw-[4-5].com	DD W

Cancel Delete

Click Delete, and the Role Types assigned to ageorge-cdsw-5.com and ageorge-cdsw-4.com hosts are deleted.

8. Delete any other active roles from the hosts.

9. In All Hosts, select the ageorge-cdsw-5.com and ageorge-cdsw-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-cdsw-[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.

Installing CML

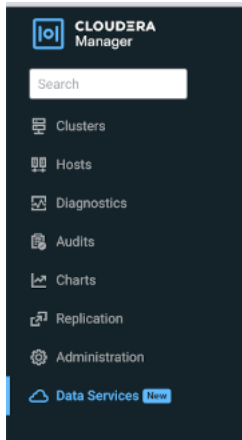
After you install Cloudera Data Platform (CDP) Private Cloud, you install Cloudera Machine Learning (CML) using UI-based tools for migration of CDSW to CML.

Before you begin

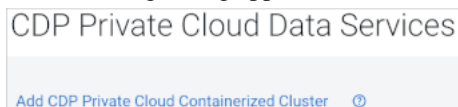
Install CDP Private Cloud Data Services.

Procedure

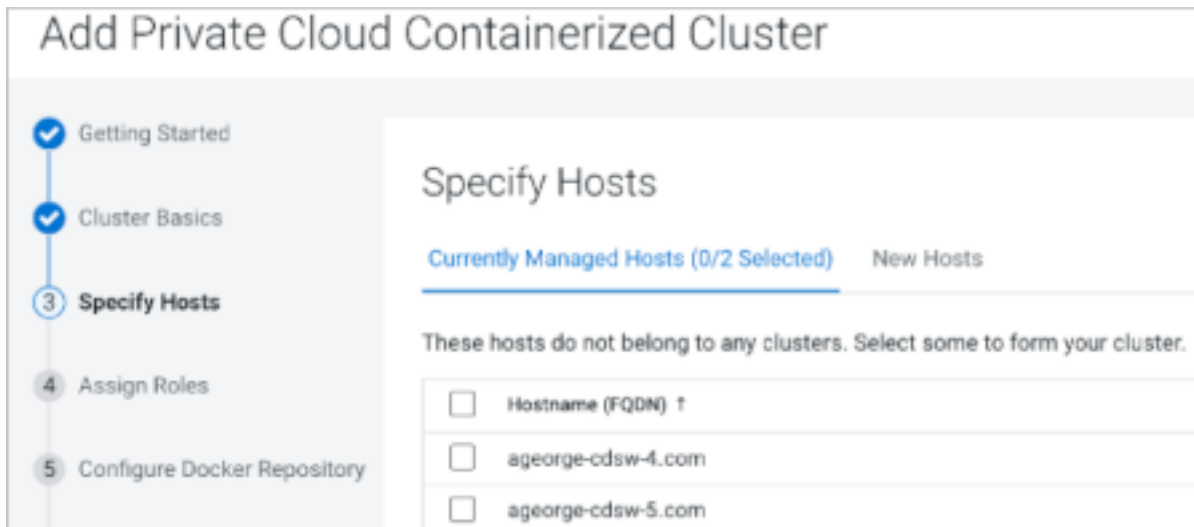
1. In Cloudera Manager, click Data Services.



The following dialog appears.



2. Click Add CDP Private Cloud Containerized Cluster.



3. In Add Private Cloud Containerized Cluster, specify the repurposed hosts.


4. In Assign Roles, complete the assignment of roles to the hosts.

For example assign the Docker Server and ECS Agent roles to ageorge-cds-5.com and the ECS role to ageorge-cds-4.com.

Assign Roles


You can customize the role assignments for your new cluster here, selected a specific host for a specific role.

You can also view the role assignments by host. [View By Host](#)

 **DOCKER**

Docker Server × 1 New

ageorge-cds-5.com

 **ECS**


Ecs Server × 1 New


ageorge-cds-4.com


Ecs Agent × 1 New

ageorge-cds-5.com


The CML Private Cloud cluster is installed, and you can start the migration from legacy CDSW to CML.

 **CLOUDERA**
Machine Learning


 **Workspaces**

 Model Registries

Machine Learning Workspaces

 **Legacy CDSW Cluster Detected**

These legacy CDSW clusters were detected:

cdsw  [Upgrade](#)

It is recommended that you upgrade your clusters to new Private Cloud instances.

Customizing CDSW for migrating host mounts

For security reasons, Cloudera Machine Learning (CML) 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 CML 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 CML. In the pre-migration steps below, you set up CDSW for the migration to mount your libraries in CML.

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 CML using the [custom runtime addons](#) procedure after migration to mount data in all the workloads in CML. 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 CML projects using that custom runtime.

Using the CDSW to CML Migration tool

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

About this task

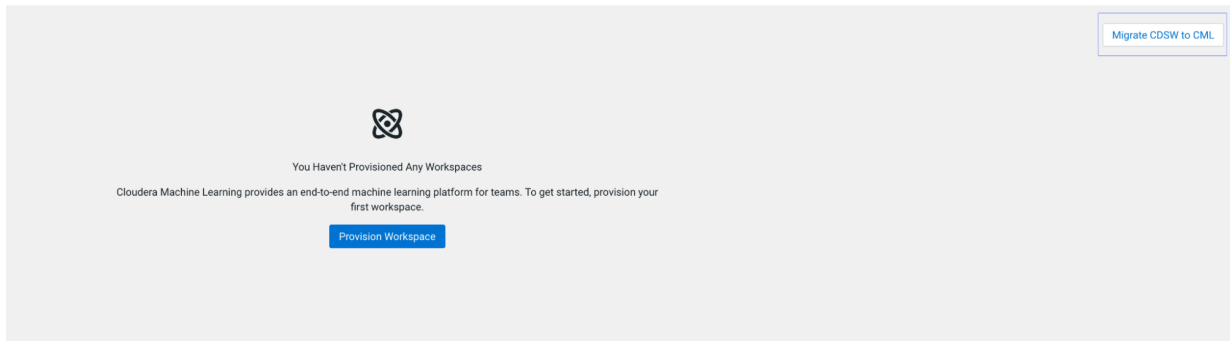
In this task, you automatically migrate CDSW 1.10.0 or later cluster to CML in Cloudera Private Cloud 1.5.1 or later. You can expect some downtime, which is proportional to the volume the workloads you have to migrate.

Before you begin

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

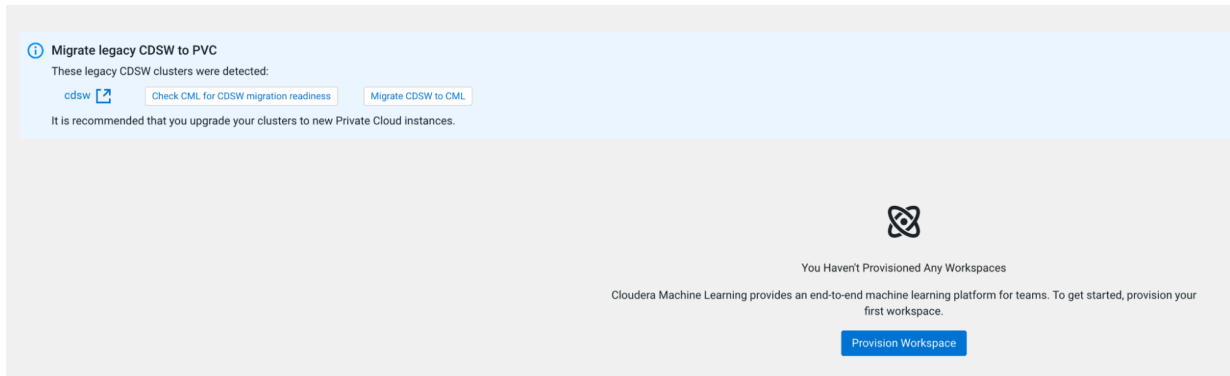
Procedure

1. Log into CDP Private Cloud 1.5.1, and navigate to Cloudera Machine Learning Workspaces .
The system detects the presence of your legacy CDSW installation and provides a button to migrate CDSW to CML.



2. Click Migrate CDSW to CML.

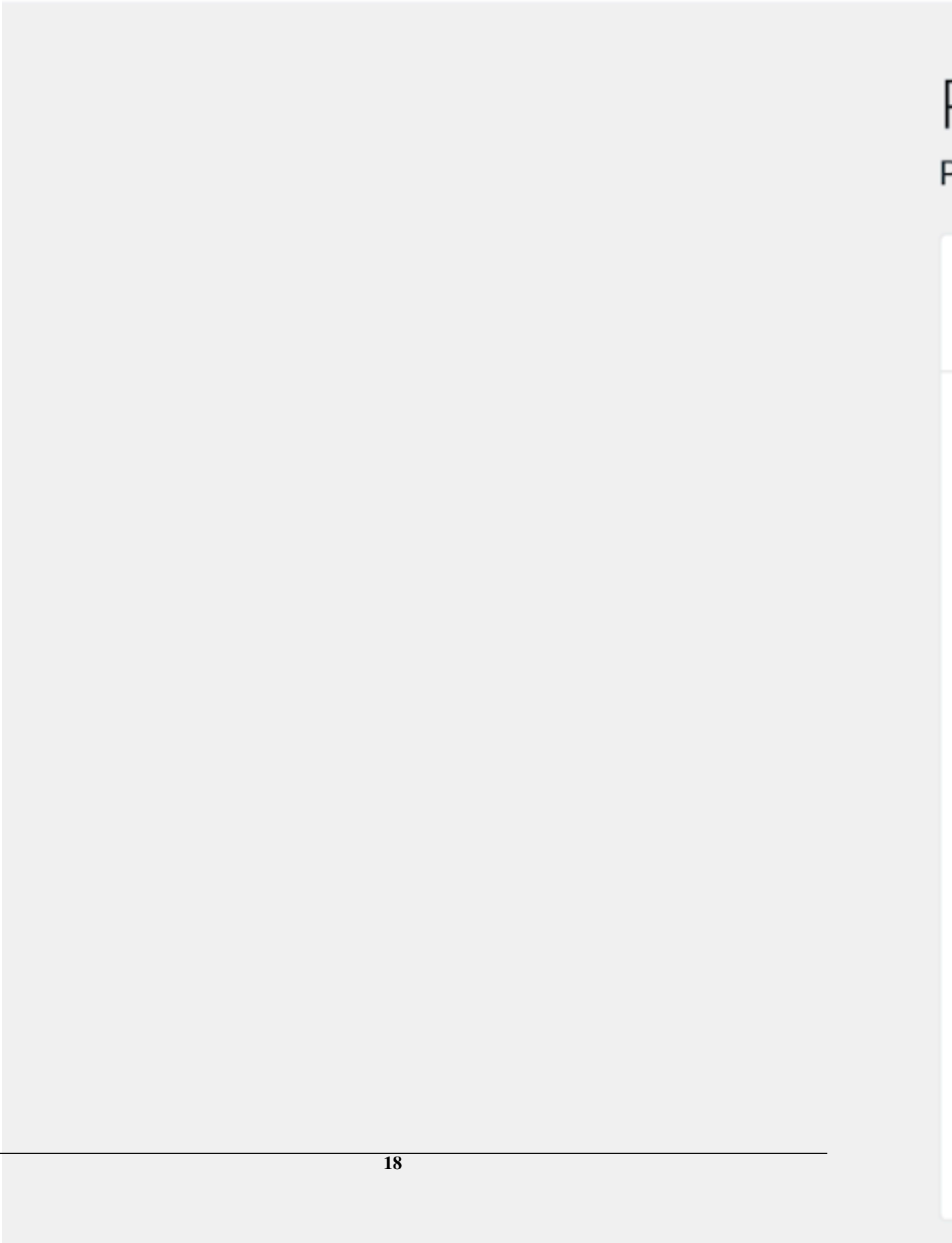
The Migration tool provides an option to check CML for CDSW migration readiness or just continue with the migration.



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

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

Validate CML for CDSW migration ...



- 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 Workspace Name field, type an arbitrary name.


- d) In the Select Environment field, select your CDP environment.

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

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

Status	Version
 Migration Readiness Failed	2.0.38

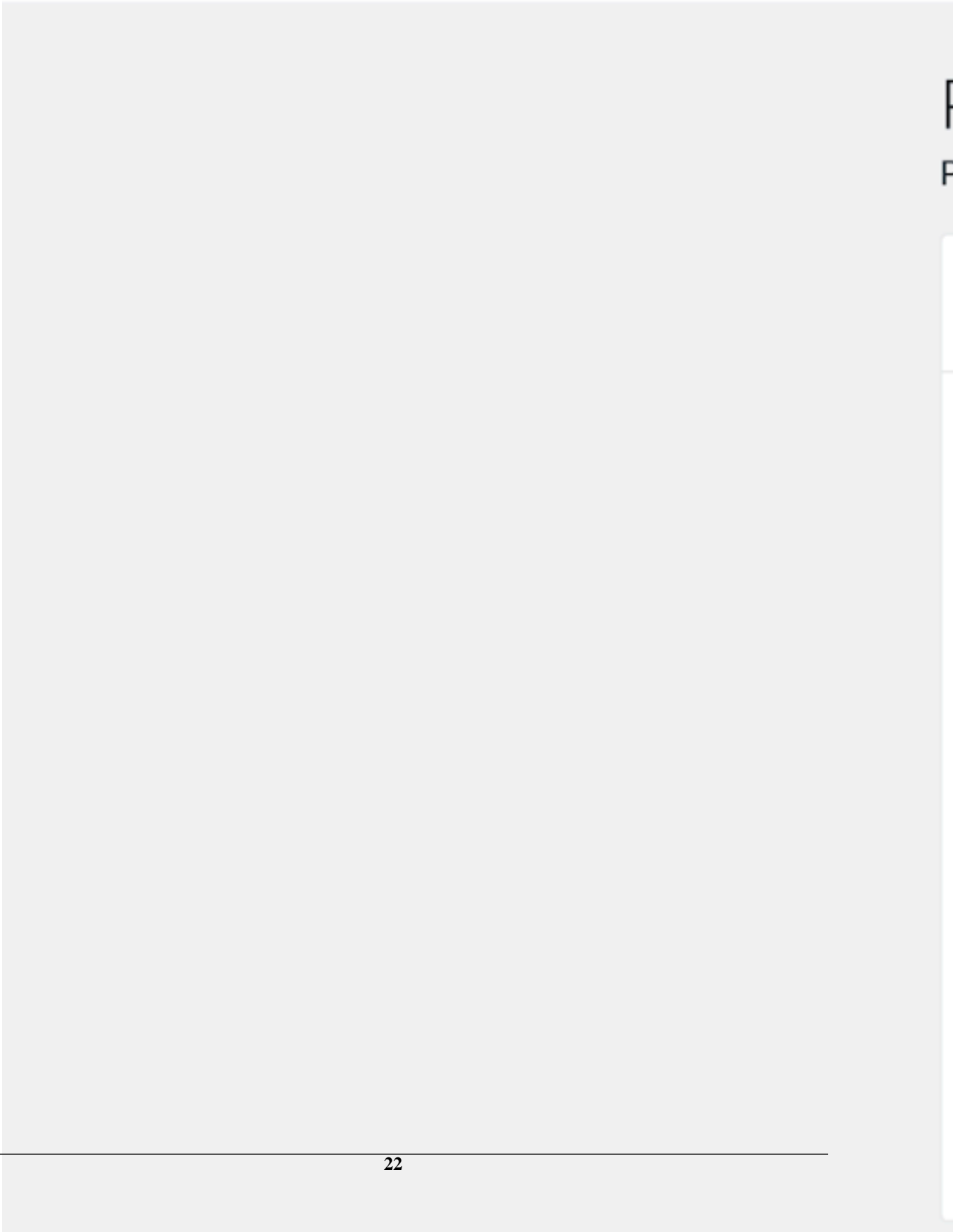
- After you've addressed the issue resulting in the failed readiness check, you can retry the readiness check by choosing Retry Migration Readiness check from the Actions menu.

Status	Version
 Migration Readiness Failed	2.0.38

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

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

Validate CML for CDSW migration ...



6. 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.

7. 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.

8. In the Workspace Name field, type an arbitrary name.
9. In the Select Environment field, select your CDP environment.

10. Accept default values for other options, and click Provision Workspace.

After the CML 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 CML workspace is stopped.



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

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

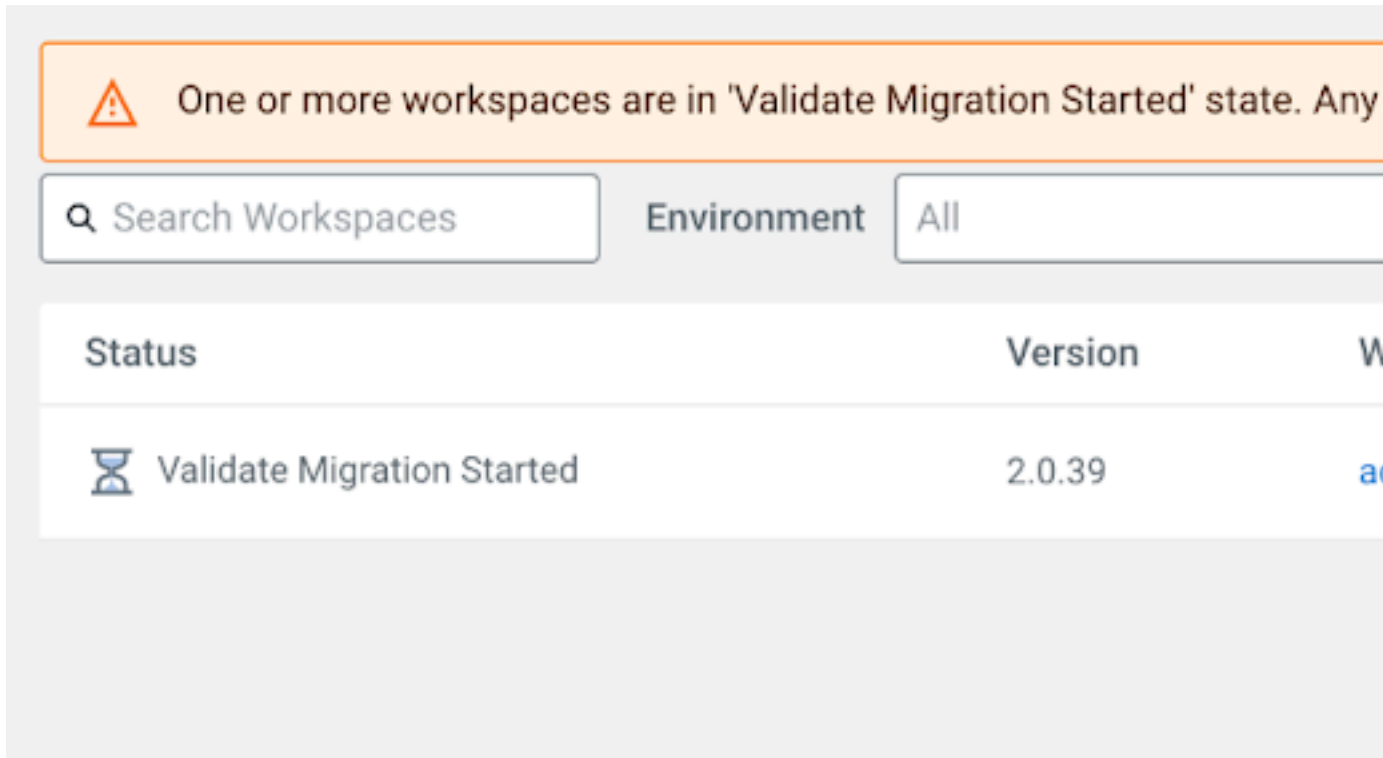
You can also view details of the migration and events and logs by clicking the appropriate tabs.

Machine Learning Workspaces / migrate

The screenshot displays the 'Machine Learning Workspaces / migrate' interface. At the top, a 'Status' box shows a blue circle icon and the text 'Migration Started'. Below this, three tabs are visible: 'Details', 'Events & Logs', and 'Migration Progress', with the latter being the active tab. The main content area is titled 'Migration Progress' and lists the migration steps under the heading 'CDSW to CML migration'. The steps are as follows:

- Running preliminary sanity checks (Completed, green checkmark)
- Verifying kubernetes configuration file (Completed, green checkmark)
- Running preflight check (Completed, green checkmark)
- Checking the migrations (Completed, green checkmark)
- Checking the auth type (Completed, green checkmark)
- Getting docker credentials (Completed, green checkmark)
- Scale down CDSW deployments (In Progress, grey circle with 'PENDING')
- Wait for database affecting jobs (Pending, blue circle with 'PENDING')
- Scaledown CML deployments (Pending, blue hourglass icon)

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



13. Open the workspace by clicking the workspace name and validate the workloads.

Any changes made to the workspace while you are validating the workspace 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 CML workspace will be active.

 One or more workspaces are in 'Validate Migration Started' state. Any

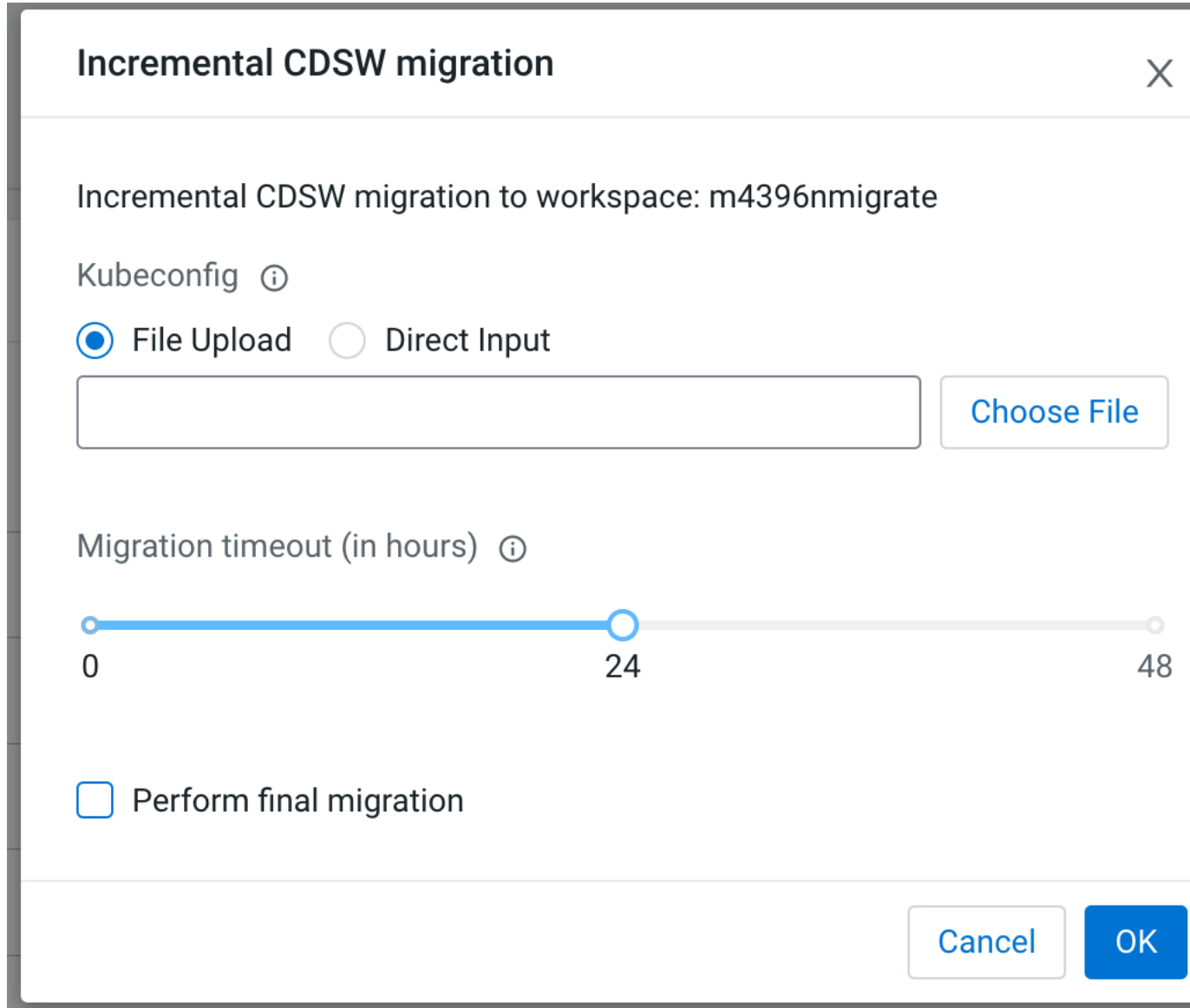
 Search Workspaces

Environment

All

Status	Version	
 Validate Migration Started	2.0.39	a

The Migration tool displays the Incremental CDSW migration dialog box.



The dialog box is titled "Incremental CDSW migration" and has a close button (X) in the top right corner. Below the title bar, the text "Incremental CDSW migration to workspace: m4396nmigrate" is displayed. Underneath, the label "Kubeconfig" is followed by an information icon (i). There are two radio buttons: "File Upload" (which is selected) and "Direct Input". Below the radio buttons is a text input field and a "Choose File" button. Further down, the label "Migration timeout (in hours)" is followed by an information icon (i). Below this is a horizontal slider with markers at 0, 24, and 48. The slider is currently set to 24. At the bottom, there is a checkbox labeled "Perform final migration". In the bottom right corner, there are two buttons: "Cancel" and "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 CML 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 workspace: 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 CML workspace. In this case, source CDSW workspace will be turned off after migration and workloads on migrated CML workspace 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 CML 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 CML 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 CML 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 CML.

Troubleshooting Preflight Migration Check Issues

Provides a 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 Machine Learning (CML) we are using core DNS. If the customers add any custom KubeDNS configurations in the CDSW, those will not be copied to CML 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 CML.

Registry Readiness

This preflight check will verify the registry readiness in CML. It will check if the created CML 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 CML. 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 CML. 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 CML. 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

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 CML. Please ensure enough storage is in the NFS storage to do the migration.

Runtime Addons

CML 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 CML properly.

Troubleshooting failures:

In CML, 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 do a reload.

Site Administration / Runtime

[Overview](#)[Users](#)[Teams](#)[Usage](#)[Quotas](#)[More](#)**Hadoop CLI Version**

Hadoop CLI - CDP 7.1.8

Runtime Addons

Status ▴ ▾	Name ▴ ▾	ID
✓ Available	Hadoop CLI - CDP 7.1.7-2000	2
✓ Available	Hadoop CLI - CDP 7.1.8	1
✓ Available	Ozone - 718.2.0-b38	3
✓ Available	Spark 2.4.7 - CDP 7.1.7.2000	4
✓ Available	Spark 3.2.3 - CDP 7.1.7.2000	5

Service Readiness

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

Troubleshooting failures:

Ensure the necessary services are started in CML during the workspace provisioning.

Versions Readiness

CDSW to CML migration 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 CDSW to CML migration

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 later to Cloudera Machine Learning (CML):

- CDSW to CML migration is not supported for the Embedded Container Service (ECS) cluster installed with the internal registry alias option.
- Migration to CDP Private Cloud 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 private cloud cluster after migration.
- CDSW and CDP Private Cloud 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 CML as the engine architecture differs between CML 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.

Post-migration tasks

After migrating Cloudera Data Science Workbench (CDSW) 1.10.0 or later to Cloudera Machine Learning (CML), you need to perform several tasks before moving users to CML.

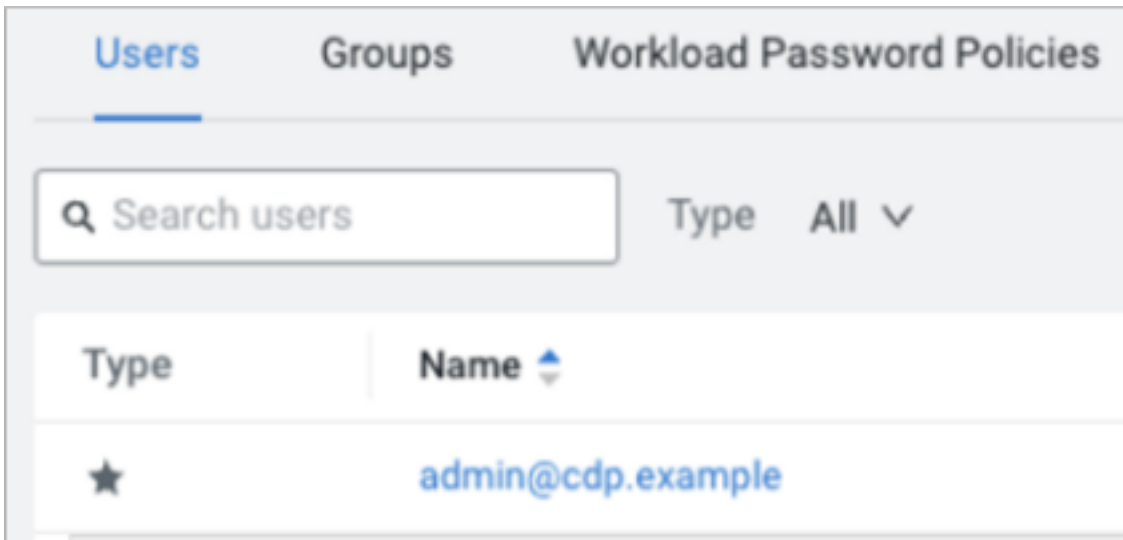
About this task

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

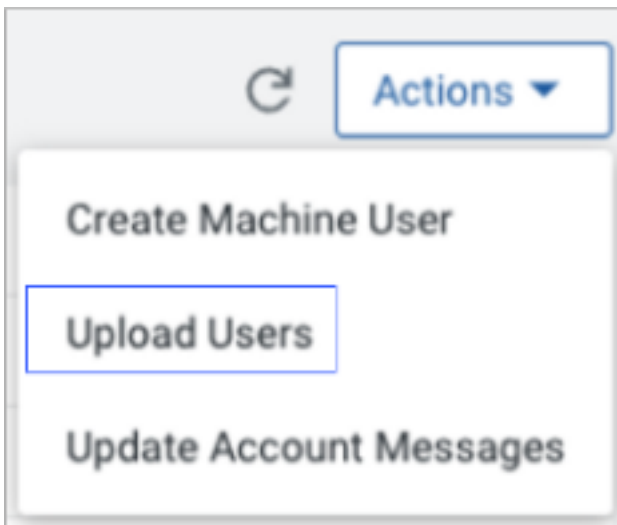
The CDSW to CML migration in Cloudera Manager updates the docker registry, engines, and model builds. You follow steps to assign roles to users and groups, import Grafana dashboards you previously exported, configure endpoints and DNS resolution, and configure LDAP.

Procedure

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



2. Click Actions Upload Users .



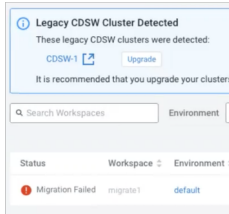
3. Select Groups, click Create Groups, and create groups.
4. Start the CML workspace on the CML cluster, and check workloads.
Start new sessions, jobs, models, and applications. Try starting existing workloads that were migrated from CDSW.
5. Import the Grafana dashboards you exported earlier.
6. Configure 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 private cloud cluster.
If you did not customize your DNS configuration, the migration tool sets up the default DNS configuration in your private cloud cluster.
8. Configure [LDAP on CML](#), and [grant user access on CML](#).
CML in CDP Private Cloud supports only LDAP.
9. Disable the CDSW cluster, and give users access to CML.

Troubleshooting CDSW to CML migration

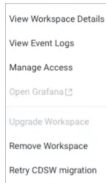
You need to know how to retry migration if your migration from Cloudera Data Science Workbench (CDSW) to Cloudera Machine Learning (CML) fails. Also, learn how to find logs for debugging migration and other issues.

Failed Migration

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



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



CDSW to CML migration frozen with large Projects

Problem: If the Project being migrated from CDSW to Cloudera Machine Learning 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 Data Platform manually that the migration has finished:


1. Find the Customer Resource Number (CRN) of the new workspace from the Control Plane.
2. Open a shell into the cdp-embedded-db-0 pod which is in the cdp 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 Data Platform 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 Workspaces, click Options , and select View Event Logs.

Problem when migrating parcels to a new CML workspace

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 CML migration script will not find them.

Workaround:

1. Check that the 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/
```