

Installing CDP Private Cloud Data Services with the Embedded Container Service

Date published: 2020-12-16

Date modified: 2022-11-18



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

CDP Private Cloud Data Services installation overview.....	4
Requirements.....	4
CDP Private Cloud Base Software Requirements.....	4
CDP Private Cloud Data Services Hardware Requirements.....	5
Requirements for HA and Non-HA Control Plane.....	5
Additional resource requirements for Cloudera Data Warehouse.....	5
Additional resource requirements for Cloudera Data Engineering.....	6
Additional resource requirements for Cloudera Machine Learning.....	7
How to use the CDP Private Cloud Data Services sizing spreadsheet.....	7
Docker repository access.....	10
CDP Private Cloud Data Services Software Requirements.....	11
Installation using the Embedded Container Service (ECS).....	11
Preparing CDP Private Cloud Base.....	11
CDP Private Cloud Base checklist.....	12
Embedded Container Service (ECS) checklist.....	14
Adding a CDP Private Cloud Data Services cluster.....	14
Installing CDP Private Cloud Data Services using ECS.....	15
ECS Server High Availability.....	29
Manually uninstalling ECS from a cluster.....	42
Upgrading.....	45
Upgrading Cloudera Manager.....	45
Update from 1.3.4 or 1.4.0-H1 to 1.4.1 (ECS).....	45

CDP Private Cloud Data Services installation overview

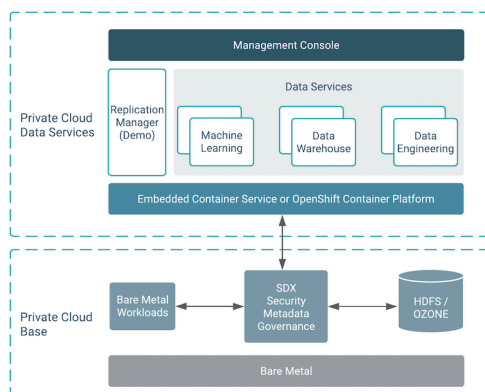
CDP Private Cloud Data Services works on top of CDP Private Cloud Base and is the on-premise offering of CDP that brings many of the benefits of the public cloud deployments to the on-premise CDP deployments. Like the cloud native applications, the CDP Private Cloud Data Services are designed to be easy to use, offer tenant-level isolation and self-service with auto-scale. All of this is made possible by the new Embedded Container Service (ECS) which manages the compute infrastructure and ease of deployment for the Data Services.



Note: To install CDP Private Cloud Data Services on OpenShift, see [Installing on OpenShift](#).

An installation of CDP Private Cloud Base is required to deploy CDP Private Cloud Data Services. A base cluster deployed in CDP Private Cloud Baseserves as a Data Lake for the Data Services. CDP Private Cloud Data Services is the on-premise offering of CDP that brings many of the benefits of public cloud deployments to on-premise CDP deployments. CDP Private Cloud Data Services allows you to deploy the Cloudera Data Warehouse (CDW), Cloudera Machine Learning (CML), and Cloudera Data Engineering(CDE) data services.

The CDP Private Cloud Data Services rely and work with customer's existing data storage and governance clusters, which we refer to as Private Cloud Base Cluster. The data services, once deployed, seamlessly and securely connect with the Private Cloud Base cluster. The following diagram is a typical example of a CDP Private Cloud Data Services deployment:



Before you can install CDP Private Cloud Data Services, you need a running instance of CDP Private Cloud Base. You need an isolated hardware environment with dedicated infrastructure and networking for CDP Private Cloud Data Services.

Requirements

CDP Private Cloud Base Software Requirements

The software requirements for the nodes on which CDP Private Cloud Data Services are deployed are identical to CDP Private Cloud Base.

Your Private Cloud Base cluster must have the operating system, JDK, database, CDP components, and CDP Runtime version compatible with CDP Private Cloud Data Services.

You can install your Private Cloud Base cluster independently. Cloudera recommends that you first set up the Private Cloud Base cluster with data and then install the Private Cloud Containerized cluster. To know the requirements for the Private Cloud Base cluster, see [Requirements and Supported Versions](#).

Ensure that you have CDP Private Cloud Base 7.1.8 or 7.1.7 SP1 with a Data Lake cluster and Cloudera Manager 7.8.1.



Note: The current Cloudera Manager version supports the Private Cloud Base cluster. You must upgrade the current Cloudera Manager to Cloudera Manager 7.8.1 to install the Private Cloud Containerized cluster.

The following CDP Private Cloud Base cluster services are required to fully access the Data Services:

- Hive Metastore (HMS)
- Ranger
- Atlas
- HDFS
- Ozone
- YARN
- Kafka
- Solr

CDP Private Cloud Data Services Hardware Requirements


Minimum and recommended hardware to successfully install and run Private Cloud Data Services.

In addition to the resources required for the Control Plane, additional resources will be required depending on the Data Service(s) you intend to run. Minimum and recommended additional resource requirements for each of the Data Services can be found in the pages below. To calculate the total minimum or recommended resource requirements for your CDP Private Cloud Data Services cluster, add the resources required for the Control Plane to the total minimum or recommended additional resources for your chosen Data Service(s).

You can also use the CDP Private Cloud Data Services Spreadsheet to model the number and specification of hosts required for a deployment. See [How to use the CDP Private Cloud Data Services sizing spreadsheet](#) on page 7.

Requirements for HA and Non-HA Control Plane

Standard resource mode requirements for standalone HA and Non-HA Control Plane.

Component	Minimum	Recommended
Node Count	1 (Non-HA)  Note: The Control Plane does not require a dedicated node.	3 (HA)
CPU	8 core	8 core (per node)
Memory	16 GB	16 GB (per node)
Storage	300 GB	1 TB (per node)
Network Bandwidth	1GB/s to all nodes and base cluster	1GB/s to all nodes and base cluster

Additional resource requirements for Cloudera Data Warehouse

Standard resource mode requirements for Cloudera Data Warehouse.

The following table lists the minimum and recommended compute (processor), memory, storage, and network bandwidth required for each OpenShift or ECS worker node using the Standard Resource Mode for production use case. Note that the actual node still needs some extra resources to run the operating system, Kubernetes engine, and Cloudera Manager agent on ECS.

Component	Minimum	Recommended
Node Count	10	20
CPU per worker	16 cores [or 8 cores or 16 threads that have Simultaneous Multithreading (SMT) enabled]	32+ cores (can also be achieved by enabling SMT)
Memory per worker	128 GB per node	384 GB* per node
FAST (Fully Automated Storage Tiering) Cache - Locally attached SCSI device(s) on every worker. Preferred: NVMe and SSD. OCP uses Local Storage Operator. ECS uses Local Path Provisioner.	1.2 TB* SATA, SSD per host	1.2 TB* NVMe/SSD per host
Persistent Volume (PV) Block Storage. On OCP, block from a Container Storage Interface (CSI)-compliant block provider, such as OpenShift Container Storage. ECS uses an embedded distributed block provider that aggregates local disks of workers. Other Data Services additionally have added capacity requirements on this service.	Approximately 100 GB per Virtual Warehouse	Approximately 100 GB per Virtual Warehouse
Network Bandwidth	1 GB/s guaranteed bandwidth to every CDP Private Cloud Base node	10 GB/s guaranteed bandwidth to every CDP Private Cloud Base node



Important: When you add memory and storage, it is very important that you add it in the increments as follows:

- Increments of 128 GB of memory
- Increments of 600 GB of locally attached SSD/NVMe storage
- Increments of 100 GB (in 5 chunks of 20 GB each) of persistent volume storage per Virtual Warehouse

If you add memory or storage that is not in the above increments, the memory and storage that exceeds these increments is not used for executor pods. Instead, the extra memory and storage can be used by other pods that require fewer resources.

For example, if you add 200 GB of memory, only 128 GB is used by the executor pods. If you add 2 TB of locally attached storage, only 1.8 TB is used by the executor pods.

Additional resource requirements for Cloudera Data Engineering

Standard resource mode requirements for standalone Cloudera Data Engineering.

Component	Minimum	Recommended
Node Count	3	5
CPU	16 cores for CDE workspace (base and virtual cluster) and 8 cores for workload	16 cores for CDE workspace (base and virtual cluster) and 32 cores (you can extend this depending upon the workload size)
Memory	64 GB for CDE workspace (base and virtual cluster) and 32 GB (you can extend this depending upon the workload size)	64 GB for CDE workspace (base and virtual cluster) and 64 GB (you can extend this depending upon the workload size)
Storage	200 GB block storage and 500 GB NFS storage	200 GB block storage and 500 GB NFS storage

* Depending on the number of executors you want to run on each physical node, the per-node requirements change proportionally. For example, if you are running 3 executor pods per physical node, you require 384 GB of memory and approximately 1.8TB (600GB per executor) of locally attached SSD/NVMe storage for FAST Cache.

Component	Minimum	Recommended
Network Bandwidth	1 GB/s to all nodes and base cluster	10 GB/s to all nodes and base cluster

Additional resource requirements for Cloudera Machine Learning

Standard resource mode requirements for standalone Cloudera Machine Learning. Node count should not be a limiting factor assuming the other MEM and CPU mins are reached.

Component	Minimum	Recommended
Node Count	1	1 per workspace + additional nodes depending on expected user workloads
CPU	16 Cores Per Workspace+ additional Cores depending on expected user workloads	32 Cores Per workspace + additional Cores depending on expected user workloads
Memory	32 GB + additional memory depending on the expected workloads	64 GB Per Workspace + additional memory depending on the expected workloads
Storage	600 GB Block storage + 1000 GB NFS storage (Block if internal and NFS if external)	4500 GB Block storage+ 1000 GB NFS storage if external
Network Bandwidth	1GB/s to all nodes and base cluster	1GB/s to all nodes and base cluster

Additional Resources for User Workloads:

Component	Minimum	Recommended
CPU	1 Core per concurrent workload	2–16 cores per concurrent workload (dependent on use cases)
Memory	2 GB per concurrent workload	4–64 GB per concurrent workload (dependent on use cases)

How to use the CDP Private Cloud Data Services sizing spreadsheet

You can use the sizing spreadsheet to model the hardware requirements for a CDP Private Cloud Data Services deployment.

Overview

The CDP Private Cloud Data Services Sizing spreadsheet is a spreadsheet that you can use to model the quantity and specifications for worker hosts required in a CDP Private Cloud Data Services deployment.

This spreadsheet is intended to use information about workloads you are planning to run and hardware specifications for worker nodes to arrive at an approximate number of worker nodes required for your deployment. Due to the complexity of estimating workloads, Cloudera recommends you review any sizing or purchasing decisions with Cloudera Professional Services before committing to those decisions.

How to access the spreadsheet

You can access the spreadsheet here: [CDP Private Cloud Data Services Sizing](#). The file is in Microsoft Excel format. You can open the file in Excel, or upload it to Google Sheets.

There are three tabs in the spreadsheet. You will make your inputs only on the Worker Node Totals tab. Do not modify the following tabs (these tabs contain data used to calculate values in the spreadsheet and should not be modified):

- Component Lookup
- K8s Resources



Important: Do not modify any cells except for the ones indicated below. Modifying the formulas in other cells will result in inaccurate calculations.

Workload inputs

The spreadsheet calculates the total amount vcores, RAM, and storage required based on information you enter about the combined workloads you intend to deploy. Then based on the hardware specifications entered, calculates the number of worker nodes required, which is displayed in cell E25.

The following sections describe values you must enter into the spreadsheet. Values are required for each Data Service you intend to deploy, and values to enter for the hardware specifications for your worker nodes.

Control plane and monitoring

Label	Cell	Description
PvC Control Plane	B3	1 required
– Monitoring	B4	Increment this number by one for each environment.

Cloudera Data Warehouse (CDW)

If you will deploy CDW, on the Worker Node Totals tab, enter the following information:

Label	Cell	Description
CDW Data Catalog (min 1 per env)	B6	Enter the number of Data Catalogs you will need in your deployment. You must have at least one Data Catalog.
CDW LLAP warehouses	B7	Enter the number of LLAP warehouses you will need for each Virtual Warehouse in your deployment.
-- LLAP Executors	B8	Enter the total number of LLAP Executors you will need in your deployment.
CDW Impala warehouses	B9	Enter the number of CDW Impala warehouses for each Virtual Warehouse you will need in your deployment.
-- Impala Coordinators (2 x for HA)	B10	Enter the number of Impala Warehouses you will need in your deployment. If you have enabled high availability, enter twice the number of Warehouses.
-- Impala Executors	B11	Enter the number of Impala Executors you will need in your deployment.
-- CDW Data Cache	B12	Enter the amount of CDW Cache space for each coordinator and executor (Default 600)
Data Viz - small	B13	Enter the size selected when creating a Data Visualization instance.
Data Viz -medium	B14	
Data Viz -large	B15	

For more information about sizing Cloudera Data Warehouse deployments, see:

- (OCP) [CDE hardware requirements](#).
- (ECS) [Additional resource requirements for Cloudera Data Engineering](#)

Cloudera Machine Learning (CML)

Sizing for a CML deployment depends on the number of concurrent jobs you expect to run and the number of Workspaces you provision.

Label	Cell	Description
CML Workspace (min of 1)	B17	Enter the number of workspaces you need in your deployment.
-- CML Small session	B18	Enter the number of concurrent small-sized sessions you intend to run.
-- CML Average session	B19	Enter the number of concurrent average-sized sessions you intend to run.

For more information about sizing the Cloudera Data Engineering service, see the following topics:

- [Additional resource requirements for Cloudera Machine Learning](#).
- (OCP) [Cloudera Machine Learning requirements](#)
- (ECS)[Cloudera Machine Learning requirements](#)x

Cloudera Data Engineering (CDE)

Label	Cell	Description
CDE Service (min/max 1 per cluster)	B21	Enter the number of CDE clusters you will need in your deployment.
CDE Virtual Cluster	B22	Enter the number of CDE Virtual Clusters you will need in your deployment.
-- CDE Small jobs	B23	Enter the number of concurrent small-sized jobs you intend to run.
-- CDE Avg Jobs	B24	Enter the number of concurrent average-sized jobs you intend to run.

For more information about sizing the Cloudera Data Engineering service, see [Additional resource requirements for Cloudera Data Engineering](#).

Worker node hardware specifications

Based on the inputs you supplied for your workloads, the spreadsheet totals the number of vcores, RAM, and storage required for the cluster in cells C20-C26. Then, based on the worker node hardware specifications you enter in cells B26-B29, divides the totals for vcores, RAM and storage by each of the worker node specifications to arrive at the required number of nodes for vcores, RAM and storage shown in cells D5-D29. The final number, in cell E27 chooses the higher value of these cells.

You may notice that the calculated values in cells D26 and D27 are different. This indicates that some nodes are oversubscribed for RAM or vcores. Adjust the hardware specifications for CPU and RAM until the two cells are closer together in value. Changing these values may also change the calculated number of worker nodes.

Label	Cell	Description
CPU recommend 32+ cores (64vcores)	B28	Enter the number of vcores for each worker node.
RAM (GB) recommend 384GB RAM	B29	Enter the amount of RAM, in gigabytes, for each worker node.
Disk (GB) Block (OCP CSI block, ECS Longhorn)	B30	Enter the number of gigabytes Block required for: - OpenShift Container Platform: CSI block - Embedded Container Service: ECS Longhorn
Disk (GB) Fast Cache for CDW (nvme,ssd)	B31	Enter the number of gigabytes of Fast Cache used in Cloudera Data Warehouse.
NFS (GB) (choose 1 from below)	B33	Enter required storage in either cell B30 or cell B31:

Label	Cell	Description
-- Embedded nfs - (subtract from Block provider) non-prod	B33	Enter the number of gigabytes storage for an embedded NFS.
-- External nfs	B35	Enter the number of gigabytes of storage for an External NFS.
ECS Master Node requires 1 for non HA - 3 for HA If you are using the Embedded Container Service, you will also need to provision a host for the ECS Master Node (a node running the ECS Server component). The values described here contain Cloudera's recommendations for specifications for the ECS Master node.	B38	Minimum: 8 vcores Recommended: 16 vcores
	B39	Minimum 16 GB RAM Recommended: 32 GB RAM
	B40	Minimum: 300 GB HDD (This amount is adequate for proof-of-concept cluster.) Recommended: 1 TB HDD

Docker repository access

You must ensure that the cluster has access to the Docker Container Repository in order to retrieve the container images for deployment.

There are several types of Docker Repositories you can use:

Embedded Repository

During installation, a Docker daemon is provisioned to act as the Repository. Passwords and certificates are auto generated. No additional set up is needed. Images are copied to the repository during installation. During upgrades, only the new and changed images are copied. Copying images generally takes one to two hours.

It is important to note that the Embedded Repository can be a single point of failure. If the node that runs the Docker Repository fails or becomes unavailable, some cluster functionalities might become unavailable. Moving the Docker Repository to another node is a complex process and will require engaging Cloudera Professional Services.

Cloudera Repository

Using the Cloudera Repository requires that the cluster have internet connectivity to the Cloudera public repository. Using the Cloudera Repository is the fastest option.

The Cloudera-hosted Docker Repository option may increase the time required to deploy or start the services in the cluster. Cloudera generates Docker Repository credentials that are identical to your payroll credentials. Refer to your welcome letter for the credentials or use the credential generator on cloudera.com to generate credentials from your license key.

This option is best suited for proof-of-concept, non-production deployments or deployments that do not have security requirements that disallow internet access.

Custom Repository

A Custom Repository is a repository that you manage in your environment and can be Enterprise grade and highly available.

During installation and upgrade, a custom script is generated that you use to copy the images. Copying images can take 4 - 5 hours.

Only TLS-enabled custom Docker Registry is supported. Ensure that you use a TLS certificate to secure the custom Docker Registry. The TLS certificate can be self-signed, or signed by a private or public trusted Certificate Authority (CA).



Important: The Custom Docker Registry is not supported for CML in CDP Private Cloud Data Services version 1.4.x.



Important: When using an Embedded Container Service cluster, passwords must not contain the \$ character.

CDP Private Cloud Data Services Software Requirements

The software requirements for the nodes on which CDP Private Cloud Data Services are deployed are identical to CDP Private Cloud Base. The most basic requirement is the operating system and JDK support.

This release ships with Cloudera Manager 7.8.1. This new version of Cloudera Manager has the support to create and manage the ECS cluster. If you have an existing Base cluster setup, managed by Cloudera Manager, you must first upgrade Cloudera Manager to version 7.8.1.

For this release, the ECS nodes will support:

- CentOS 8.4, 7.9, Red Hat Enterprise Linux 8.4, 7.9, Oracle Linux 7.9, and CentOS 8.2 (CDW only)
- JDK 11 (any distribution)
- For CML, you must install `nfs-utils` in order to mount longhorn-nfs provisioned mounts. The `nfs-utils` package is required on every node of the ECS cluster. Run this command `yum install nfs-utils` to install `nfs-utils`.
- If you have nodes with GPU, ensure that the GPU hosts have `nVidia Drivers` and `nvidia-container-runtime` installed. You must confirm that drivers are properly loaded on the host by executing the command `nvidia-smi`. You must also install the `nvidia-container-toolkit` package.

Additionally, you must perform the following:

- You must have a minimum of one agent node for ECS.
- Set up Kerberos on these clusters using an Active Directory or MIT KDC.
- Enable TLS on the Cloudera Manager cluster for communication with components and services.
- Configure PostgreSQL database as an external database for the Private Cloud Base cluster components*.
- If you do not have entitlements, contact your Cloudera account team to get the necessary entitlements.
- The default docker service uses `/docker` folder. Whether you wish to retain `/docker` or override `/docker` with any other folder, you must have a minimum of 200 GB space.
- Ensure all the hosts in the ECS cluster have 100 gigabytes free in the `/var/lib` directory at the time of installation.
- The cluster generates multiple hosts and host based routing is used in the cluster in order to route it to the right service. You must decide on a domain for the services which Cloudera Manager by default points to one of the hostnames on the cluster. However, during the installation, you should check the default domain and override the default domain (only if necessary) with what you plan to use as the domain. The default domain must have a wildcard DNS entry. For example, `*.apps.myhostname.com`.
- It is recommended that you leave IPv6 enabled at the OS level on all ECS nodes.

Installation using the Embedded Container Service (ECS)

Preparing CDP Private Cloud Base

Use Cloudera Manager to configure your Private Cloud Base cluster in preparation for the Private Cloud Data Services installation.

1. Configure the Private Cloud Base cluster to use TLS. [Configuring TLS Encryption for Cloudera Manager Using Auto-TLS](#).

* Cloudera Data Warehouse (CDW) supports MariaDB, MySQL, PostgreSQL versions 12 and 10, and Oracle databases for the Hive Metastore (HMS) on the base CDP cluster (Cloudera Manager side).

2. Configure Cloudera Manager with a JKS-format (not PKCS12) TLS truststore.[Database requirements](#).
3. Configure Cloudera Manager to include a root certificate that trusts the certificate for all Cloudera Manager server hosts expected to be used with the Private Cloud, LDAP server (if you are using LDAP), and the Postgres DB of all Hive Metastores that you use with Private Cloud. If a single CA is used to sign all of them, then just that single CA must be imported.
 - a. Import the necessary certificates into the truststore configured in Configure Administration > Settings > Security > Cloudera Manager TLS/SSL Client Trust Store File.
4. Enable Kerberos for all the services in the cluster.[Enabling Kerberos for authentication](#).
5. Configure Ranger and LDAP for user authentication. Ensure that you have configured Ranger user synchronization.[Configure Ranger authentication for LDAP](#) and [Ranger usersync](#).
6. Configure LDAP using Cloudera Manager. Only Microsoft Active Directory (AD) and OpenLDAP are currently supported.[Configure authentication using an LDAP-compliant identity service](#).
7. Check if all the running services in the cluster are healthy. To check this using Cloudera Manager, go to Cloudera Manager > Clusters > [***CLUSTER NAME***] > Health Issues. If there are no health issues, the No Health Issues message is displayed.
8. If you want to reuse data from your legacy CDH or HDP deployment in your Private Cloud, copy the data from your CDH or HDP deployments into the CDP Private Cloud Base cluster that will be accessed by CDP Private Cloud Data Services. For more information about data migration, see the [Data Migration Guide](#).
9. For installing CDP Private Cloud Base, see [Install CDP Private Cloud Base](#)

CDP Private Cloud Base checklist

Use this checklist to ensure that your CDP Private Cloud Base is configured and ready for installing CDP Private Cloud Data Services.



Note: The Cloudera Manager mentioned in this checklist is the CDP Private Cloud Base Cloudera Manager using which you want to install CDP Private Cloud Data Services.

Table 1: CDP Private Cloud Base checklist to install CDP Private Cloud Data Services

Item	Summary	Documentation	Notes
Runtime version and components	Ensure that you have CDP Private Cloud Base 7.1.8 or 7.1.7 SP1 with a Data Lake cluster (CDP Private Cloud Base 7.1.7 SP1 is compatible with CDP Private Cloud Data Services 1.4.1. if you upgrade to Cloudera Manager 7.8.1). Ensure that you have Ranger, Atlas, Hive, and Ozone installed in your CDP Private Cloud Base cluster.	CDP Private Cloud Base requirements	If you do not install these components, you see an error when creating an environment in your CDP Private Cloud Data Services.
Network requirement	Ensure that the CDP Private Cloud Base and OpenShift Container Platform are on the same network.	Network infrastructure considerations	Ensure that all the network routing hops in production. Cloudera recommends not to use more than 4:1 oversubscription between the spine-leaf switches.
Cloudera Manager database requirement	Refer to the the CDP Private Cloud Base database requirements.	<ul style="list-style-type: none"> Database Requirements Cloudera Support Matrix 	N/A
Cloudera Manager TLS configuration	Ensure that Cloudera Manager in the CDP Private Cloud Base cluster is configured to use TLS.	Configuring TLS Encryption for Cloudera Manager Using Auto-TLS	You can also manually configure TLS to complete this task. See Manually Configuring TLS Encryption for Cloudera Manager

Item	Summary	Documentation	Notes
Cloudera Manager JKS-format TLS truststore	Ensure that the Cloudera Manager is configured with a JKS-format (not PKCS12) TLS truststore.	Obtain and Deploy Keys and Certificates for TLS/SSL	N/A
Cloudera Manager truststore and root certificate	Ensure that the Cloudera Manager truststore contains a root certificate that trusts the certificate for all Cloudera Manager server hosts used with CDP Private Cloud Data Services.	How to Add Root and Intermediate CAs to Truststore for TLS/SSL	Import the necessary certificates into the truststore configured in <code>Configure Administration > Settings > Security > Cloudera Manager TLS/SSL Client Trust Store File</code> .
LDAP configuration	Ensure that you configure LDAP using Cloudera Manager.	N/A	Only Microsoft Active Directory (AD) and OpenLDAP are currently supported.
Apache Ranger configuration for LDAP	Ensure that the CDP Private Cloud Base cluster is configured with Apache Ranger and LDAP for user authentication.	Configure Ranger authentication for LDAP	N/A
Apache Ranger usersync configuration	Ensure that you have configured Apache Ranger and Apache Ranger usersync.	Ranger usersync	Apache Ranger user synchronization is used to get users and groups from the corporate ActiveDirectory to use in policy definitions.
Kerberos configuration	Ensure that Kerberos is enabled for all services in the cluster.	Enabling Kerberos for authentication	Custom Kerberos principals are not currently supported.
Internet access or air gap installation	Ensure that your CDP Private Cloud Base and your Kubernetes cluster has access to the Internet. If you do not have access to the Internet, you must do an air gap installation.	CDP Private Cloud Base requirements Install CDP Private Cloud Data Services in air gap environment	You need access to the Docker registries and the Cloudera repositories during the installation process.
Services health check	Ensure that all services running in the cluster are healthy.	Cloudera Manager Health Tests	N/A
CDP Private Cloud entitlement	Ensure that you have the necessary CDP entitlement from Cloudera to access the Private Cloud installation.	N/A	To check this, go to <code>Cloudera Manager > Private Cloud > Select Repository > [***REPOSITORY URL***]</code> . If you have the required entitlements, the <code>You are about to install the CDP Private Cloud version [***VERSION***]</code> message with a list of prerequisites is displayed. An error message is displayed if you do not have the necessary entitlements.
Reuse data from CDH or HDP (Optional)	To reuse data from your legacy CDH or HDP deployment in your Private Cloud, ensure that you have migrated that data into your CDP Private Cloud Base. You must be using Cloudera Runtime 7.1.7 for migrating data from your CDH or HDP cluster.	Data Migration Guide	N/A

Item	Summary	Documentation	Notes
(Recommended) Configure HDFS properties to optimize logging	CDP uses “out_webhdfs” Fluentd output plugin to write records into HDFS, in the form of log files, which are then used by different data services to generate diagnostic bundles. To optimize the size of logs that are captured and stored on HDFS, you must update a few HDFS configurations in the hdfs-site.xml file using Cloudera Manager.	Configuring HDFS properties to optimize logging	N/A

Embedded Container Service (ECS) checklist

Use this checklist to ensure that your Embedded Container Service (ECS) is configured and ready for installing CDP Private Cloud Data Services.

Table 2: Embedded Container Service (ECS) checklist to install CDP Private Cloud Data Services

Item	Summary	Documentation	Notes
DNS configuration	Ensure that you have set up the DNS and Reverse DNS between Embedded Container Service (ECS) hosts and CDP Private Cloud Base. This is required for obtaining Kerberos ticket-granting tickets.	Certificate management and DNS	A wildcard DNS entry is required for resolving the ingress route for applications. The ingress route is usually behind a load balancer.
Check that ECS Ingress can be resolved in DNS.	Ensure that Embedded Container Service (ECS) application hostnames can be accessed from outside the cluster. You can test this by creating an ingress point on the target cluster.	The cluster generates multiple hosts and host-based routing is used in the cluster in order to route it to the right service. You must decide on a domain for the services which Cloudera Manager, by default points to one of the hostnames on the cluster. However, during the installation, you should check the default domain and override the default domain (only if necessary) with what you plan to use as the domain. The default domain must have a wildcard DNS entry. For example, *.apps.myhostname.com.	Perform a DNS query on the ingress point, to check if you can access the hostnames outside the cluster.
Clock time from NTP source	Ensure that the NTP clock in CDP Private Cloud Base is in sync with the time configured in the Embedded Container Service (ECS) cluster. This is an important step if your setup does not have access to the Internet.	Enable an NTP Service	Installing CDP Private Cloud Data Services (ECS)

Adding a CDP Private Cloud Data Services cluster

Using the new Cloudera Manager 7.8.1, you can either install Private Cloud Data Services by downloading the repository from the Internet or you can do an air gap installation if Cloudera Manager does not have access to the Internet.

Before you begin:

- Ensure you have Cloudera Manager 7.8.1 installed and you have the entitlements to the CDP Private Cloud Data Services product.
- Only TLS 1.2 is supported for authentication with Active Directory/LDAP. You require TLS 1.2 to authenticate the CDP control plane with your LDAP directory service like Active Directory.
- When the system logs you out after some period of inactivity and the Add Cluster wizard is not yet finished, you must either restart Cloudera Manager or from the Home screen > select Add Cluster > choose CDP Private Cloud Base cluster > Click Continue, and then just click Cancel to leave the wizard.
- The Kubeconfig file is available in /etc/rancher/rke2/rke2.yaml
- If the installer fails, do not cancel the installation. For more information, see [Manually uninstalling ECS from a cluster](#).
- If you are installing ECS on RHEL 8.4, step 2 and its substeps available in the next section are mandatory.



Note: While installing the ECS cluster using the Cloudera Manager UI, do not provide the unsupported characters in the app-domain. The supported characters are:

- Lower case alphanumeric characters
- '-' or '.'
- It must start and end with an alphanumeric character. For example, 'example.com', and regex used for validation is '[a-z0-9]([-a-z0-9][a-z0-9])?(\.[-a-z0-9]([-a-z0-9][a-z0-9])?)*'

Installing CDP Private Cloud Data Services using ECS

Follow the steps in this topic to install CDP Private Cloud Data Services.

Procedure

1. If your ECS hosts are running the CentOS 8.4 or OEL 8.4 operating systems, you must install iptables on all the ECS hosts. (This step is not required when running RHEL 8.4.) Run the following command on each ECS host:

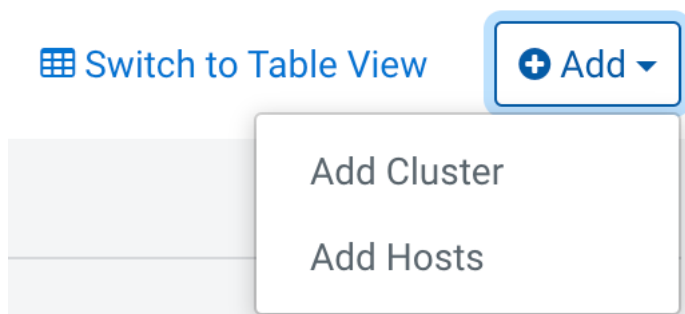
```
yum --setopt=tsflags=noscripts install -y iptables
```

2. If you are installing ECS on RHEL 8.x:,
 - a) Add the hosts you intend to use for ECS to Cloudera Manager, without specifying a cluster. See [Add New Hosts To Cloudera Manager](#).
 - b) If you are using RHEL 8.4, and if the nm-cloud-setup.service and nm-cloud-setup.timer services are enabled, disable them by running the following command on each host you added:

```
systemctl disable nm-cloud-setup.service nm-cloud-setup.timer
```

For more information, see [Known issues and limitations](#).


- c) If you disabled the nm-cloud-setup.service and nm-cloud-setup.timer services, reboot the added hosts.
3. In Cloudera Manager, on the top right corner, click Add > Add Cluster. The Select Cluster Type page appears.




4. In the Select Cluster Type page, select the cluster type as Private Cloud Containerized Cluster and click Continue.

Add Cluster

Select Cluster Type



Private Cloud Base Cluster
Add a cluster to provide storage and metadata for a compute cluster or to run workloads that benefit from data locality.



Private Cloud Containerized Cluster New
Add a Private Cloud Containerized Cluster to access our latest data analytic data services on a container cloud with separated compute and storage.
Selected

CDP Private Cloud is a next-generation data platform with container-native, self-service analytic data services bringing the speed, scale, and economics of the cloud to on-premise data centers.

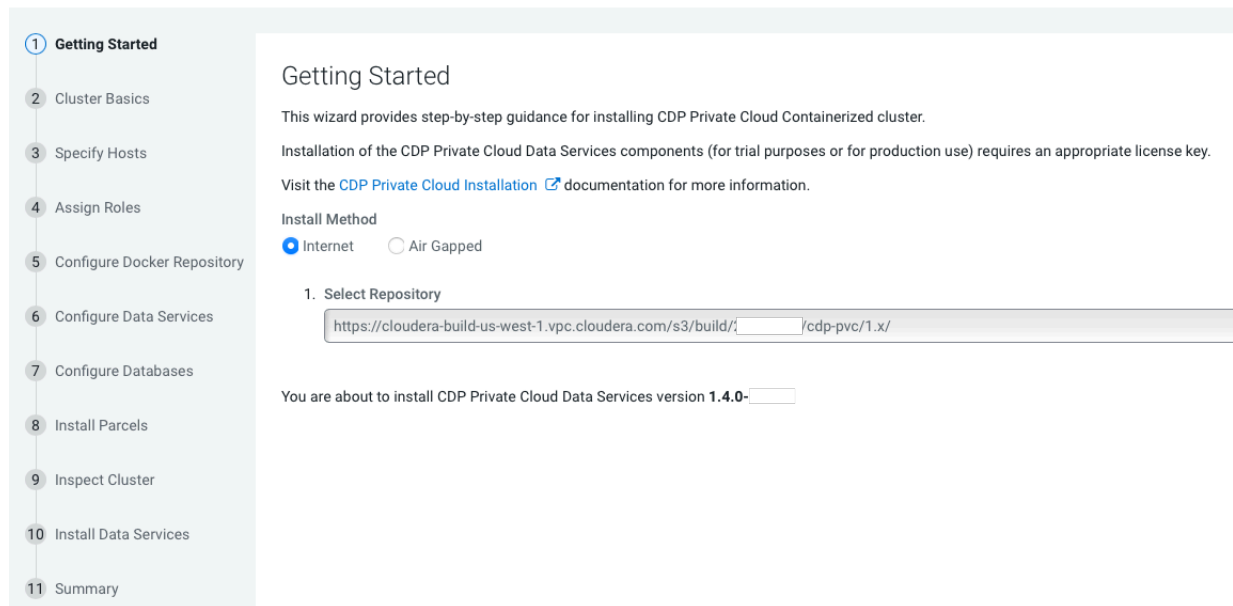
Click **Continue** to add a CDP Private Cloud Containerized Cluster, accessing data stored in HDFS or Ozone on an existing storage cluster running Cloudera Runtime 7.x. This cluster will be managed by this Cloudera Manager instance.

▼ Other Options
Click [here](#) to install the same CDP Private Cloud Data Services as above, but in a separately provisioned and managed container application platform such as OpenShift. Cloudera Manager will not be managing this OpenShift instance.

- On the Getting Started page of the installation wizard, select Internet or Air Gapped as the Install Method. To use a custom repository link provided to you by Cloudera, click Custom Repository. Click Continue.

Internet install method:

Add Private Cloud Containerized Cluster



1 Getting Started

Getting Started

This wizard provides step-by-step guidance for installing CDP Private Cloud Containerized cluster.

Installation of the CDP Private Cloud Data Services components (for trial purposes or for production use) requires an appropriate license key.

Visit the [CDP Private Cloud Installation](#) documentation for more information.

Install Method

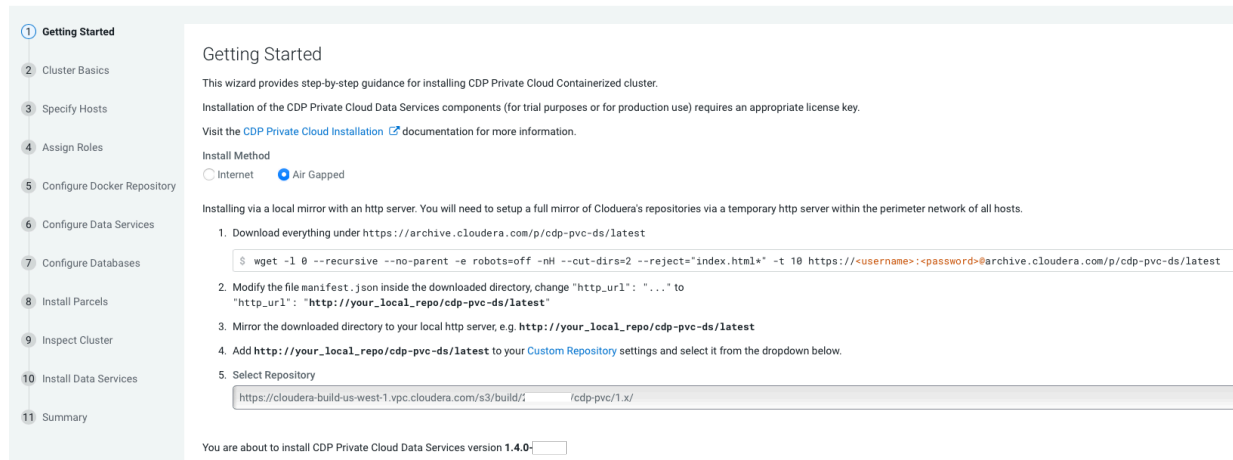
☒ Internet ☐ Air Gapped

1. Select Repository

You are about to install CDP Private Cloud Data Services version **1.4.0**.

Air Gapped install method:

Add Private Cloud Containerized Cluster



1 Getting Started

Getting Started

This wizard provides step-by-step guidance for installing CDP Private Cloud Containerized cluster.

Installation of the CDP Private Cloud Data Services components (for trial purposes or for production use) requires an appropriate license key.

Visit the [CDP Private Cloud Installation](#) documentation for more information.

Install Method

☐ Internet ☒ Air Gapped

Installing via a local mirror with an http server. You will need to setup a full mirror of Cloudera's repositories via a temporary http server within the perimeter network of all hosts.

1. Download everything under `https://archive.cloudera.com/p/cdp-pvc-ds/latest`

```
$ wget -l 0 --recursive --no-parent -e robots=off -nH --cut-dirs=2 --reject="index.html*" -t 10 https://<username>:<password>@archive.cloudera.com/p/cdp-pvc-ds/latest
```

2. Modify the file `manifest.json` inside the downloaded directory, change `"http_url": "..."` to `"http_url": "http://your_local_repo/cdp-pvc-ds/latest"`

3. Mirror the downloaded directory to your local http server, e.g. `http://your_local_repo/cdp-pvc-ds/latest`

4. Add `http://your_local_repo/cdp-pvc-ds/latest` to your [Custom Repository](#) settings and select it from the dropdown below.

5. Select Repository

You are about to install CDP Private Cloud Data Services version **1.4.0**.

Click Continue.

6. In the Cluster Basics page, type a name for the Private Cloud cluster that you want to create in the Cluster Name field. From the Base Cluster drop-down list, select the cluster that has the storage and SDX services that you want this new Private Cloud Data Services instance to connect with. Click Continue.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary

Cluster Basics

Cluster Name

Containerized Cluster 1

Private Cloud Containerized Cluster

A Private Cloud Containerized Cluster helps you to install and run CDP Private Cloud Data Services such as Machine Learning and Data Warehouse with data from an existing Base Cluster. Learn more at [CDP Private Cloud Containerized Cluster](#).

Base Cluster

Cluster 1 (7.1.7)

☐ Use Default Configuration

Use embedded Docker Repository, Vault and Database with default settings, and use default configurations for Role Assignments. Not recommended for production.

7. In the Specify Hosts page, provide a list of available hosts or you can add new hosts. (If you already added the hosts to Cloudera Manager, enter the Fully Qualified Domain Name (FQDN) for those hosts.) You can provide the FQDN in the following patterns:

You can specify multiple addresses and address ranges by separating them by commas, semicolons, tabs, or blank spaces, or by placing them on separate lines. Use this technique to make more specific searches instead of searching overly wide ranges.

For example, use host[1-3].network.com to specify these hosts: host1.network.com, host2.network.com, host3.network.com.

Click Continue.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary

Specify Hosts

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with.

Hostname

Hint: Search for hostnames or IP addresses using [pattern](#)

SSH Port

22

Search

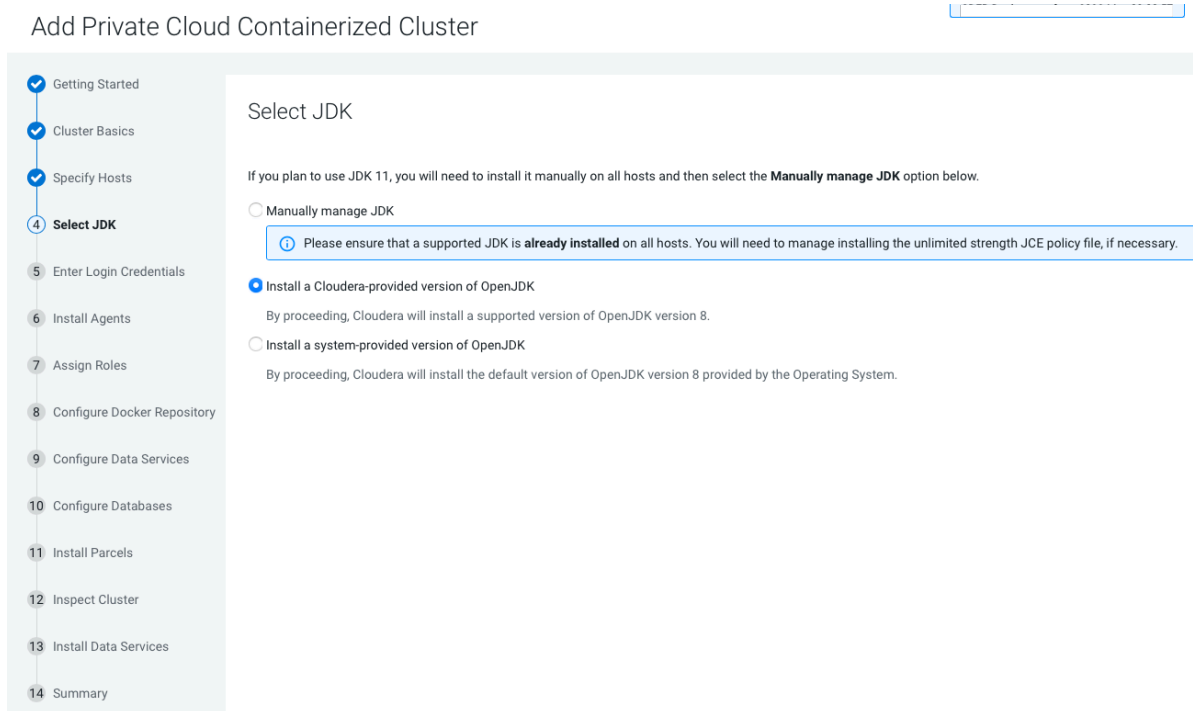
3 hosts scanned, 3 running SSH.

	Expanded Query	Hostname (FQDN) 1	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/>			10.65.6.9	No	Host was successfully scanned.
<input checked="" type="checkbox"/>			10.65.10.73	No	Host was successfully scanned.
<input checked="" type="checkbox"/>			10.65.9.254	No	Host was successfully scanned.

8. In the Select JDK page, select any one from the below options:

- a) Manually manage JDK
- b) Install a Cloudera-provided version of OpenJDK
- c) Install a system-provided version of OpenJDK

Add Private Cloud Containerized Cluster



Select JDK

If you plan to use JDK 11, you will need to install it manually on all hosts and then select the **Manually manage JDK** option below.

☐ Manually manage JDK

Please ensure that a supported JDK is **already installed on all hosts. You will need to manage installing the unlimited strength JCE policy file, if necessary.**

☒ **Install a Cloudera-provided version of OpenJDK**

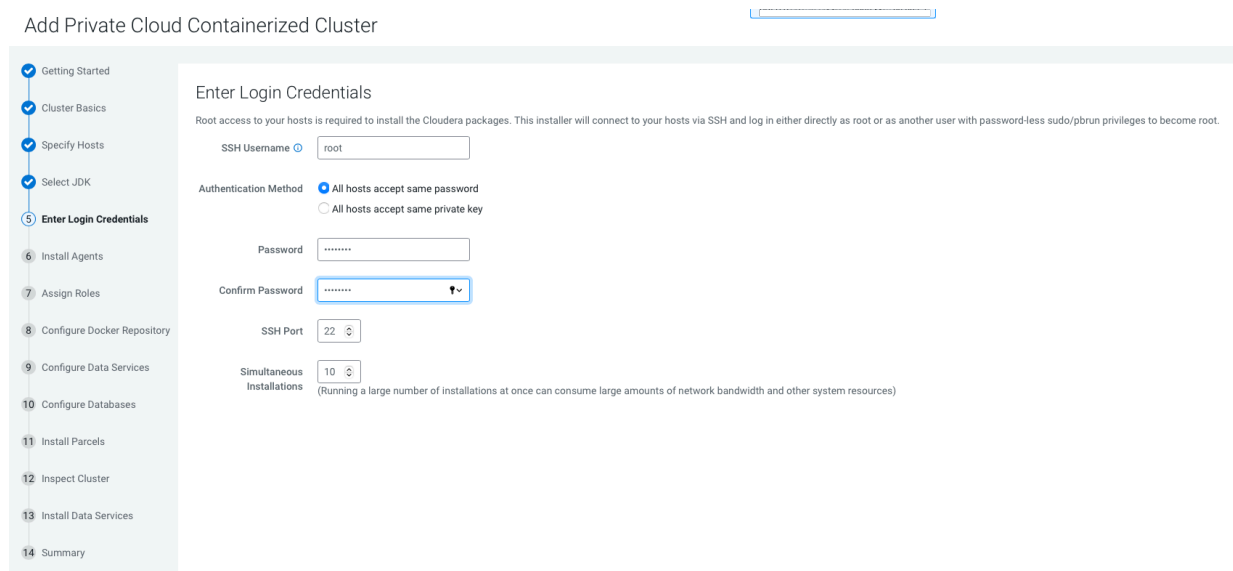
By proceeding, Cloudera will install a supported version of OpenJDK version 8.

☐ Install a system-provided version of OpenJDK

By proceeding, Cloudera will install the default version of OpenJDK version 8 provided by the Operating System.

9. In the Enter Login Credentials page select the SSH Username and provide the password.

Add Private Cloud Containerized Cluster



Enter Login Credentials

Root access to your hosts is required to install the Cloudera packages. This installer will connect to your hosts via SSH and log in either directly as root or as another user with password-less sudo/pbrun privileges to become root.

SSH Username

Authentication Method ☒ All hosts accept same password
☐ All hosts accept same private key

Password

Confirm Password

SSH Port

Simultaneous Installations
 (Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)

10. The Install Agents page appears.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary


Install Agents

Installation in progress.

0 of 3 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status	
	10.65.6.9	<div></div>	C Installing openjdk8 package...	Details
	10.65.10.73	<div></div>	C Installing openjdk8 package...	Details
	10.65.9.254	<div></div>	C Installing openjdk8 package...	Details

11. In the Assign Roles page, you can customize the roles assignment for your new Private Cloud Containerized cluster.



Important: Cloudera does not recommend altering assignments unless you have specific requirements such as having selected a specific host for a specific role.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary

Assign Roles

You can customize the role assignments for your new cluster here, but if assignments are made incorrectly, such as assigning too many roles to a single host, this can impact the performance of your services. Cloudera does not recommend altering assignments unless you have specific requirements, such as having pre-selected a specific host for a specific role. [View By Host](#)

DOCKER

Docker Server × 1 New

ECS

Ecs Server × 1 New

Ecs Agent × 2 New

Click Continue.

12. Configure a Docker Repository.

There are several options for configuring a Docker Repository. For more information about these options, see [Docker repository access](#) on page 10.

Add Private Cloud Containerized Cluster

The screenshot shows a vertical progress bar on the left with 14 steps. Steps 1 through 7 are completed, indicated by blue checkmarks. Step 8, 'Configure Docker Repository', is the current step, highlighted with a blue circle and the number 8. Steps 9 through 14 are not yet completed, indicated by grey circles with numbers. To the right of the progress bar, the title 'Configure Docker Repository' is displayed. Below the title, a paragraph states: 'Cloudera uses a Docker Repository to deliver CDP Private Cloud Data Services. [Learn more](#) about how to set up custom Docker Repository for CDP Private Cloud Data Services.' Below this paragraph, there are three radio button options: 'Use an embedded Docker Repository', 'Use Cloudera's default Docker Repository', and 'Use a custom Docker Repository'. The first option is selected.

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

8 Configure Docker Repository

9 Configure Data Services

10 Configure Databases

11 Install Parcels

12 Inspect Cluster

13 Install Data Services

14 Summary

Configure Docker Repository

Cloudera uses a Docker Repository to deliver CDP Private Cloud Data Services. [Learn more](#) about how to set up custom Docker Repository for CDP Private Cloud Data Services.

☒ Use an embedded Docker Repository

☐ Use Cloudera's default Docker Repository

☐ Use a custom Docker Repository

On the Configure Docker Repository page, select one of these options:

- Embedded Docker Repository

Add Private Cloud Containerized Cluster

Configure Docker Repository

Cloudera uses a Docker Repository to deliver CDP Private Cloud Data Services. [Learn more](#) about how to set up custom Docker Repository for CDP Private Cloud Data Services.

☒ Use an embedded Docker Repository
☐ Use Cloudera's default Docker Repository
☐ Use a custom Docker Repository

This release comes with 238 container images that need to be deployed to the Docker repository. Some images are optional and can be skipped by toggling them from the list below. Other images are always installed.

☐ Default ☒ Select the Optional Images

☒ Cloudera Machine Learning
 Docker images required to create a Cloudera Machine Learning workspace. Without these images, it will not be possible to use Cloudera Machine Learning.

The system will deploy 238 container images, approximately 82.7 GiB, to the embedded Docker repository.

If you select the Internet Install Method option on the Getting Started page, images are copied over the internet from the Cloudera repository.

If you select the Air Gapped option, images are copied from a local http mirror you have set up in your environment.

Select **Default** to deploy all of the default Docker images to the repository, or select **Select the Optional Images** to choose which images to deploy. If you will be deploying Cloudera Machine Learning (CML), toggle the Cloudera Machine Learning switch on to copy the images for CML.

- Cloudera default Docker Repository

This option requires that cluster hosts have access to the internet and you have selected Internet as the install method.

- a. Ensure that the following ports are opened and allowed. This is required for completing the ECS installation.

Protocol	Port
TCP	7180-7192
TCP	19001
TCP	5000
TCP	9000

- b. Inbound rules for ECS Server nodes.

Protocol	Port
TCP	9345
TCP	6443
UDP	8472
TCP	10250
TCP	2379

Protocol	Port
TCP	2380
TCP	30000-32767

c. Inbound Rules for ECS Agent.

Protocol	Port
UDP	4789

• Custom Docker Repository

Add Private Cloud Containerized Cluster

Configure Docker Repository

Cloudera uses a Docker Repository to deliver CDP Private Cloud Data Services. [Learn more](#) about how to set up custom Docker Repository for CDP Private Cloud Data Services.

☐ Use an embedded Docker Repository
☐ Use Cloudera's default Docker Repository
☒ Use a custom Docker Repository

Custom Docker Repository [?](#)

Prepare your Docker Repository from a machine that is running Docker locally and has access to all the Docker images either directly from Cloudera or from a local http mirror in your network. If your custom repository already has all the Docker images for this version, this section can be skipped.

1. [Generate the copy-docker script](#)
2. Optionally, review the script. The file contains usage information and lists the Docker images that it will download and push.
3. Login to your custom Docker Registry and run the script with the following commands (Note: this downloads 100+ Docker images and it will take a while):

```
docker login <your_custom_registry> -u <user_with_write_access>
bash copy-docker.txt
```

☐ I confirm that I have downloaded all the Docker images to my custom Docker Repository.

Docker Username [?](#)

Docker Password [?](#)

Docker Certificate [?](#)

[Choose File](#)

This option requires that you set up a Docker Repository in your environment and that all cluster hosts have connectivity to the repository.

You must enter the following options:

- Custom Docker Repository – Enter the URL for your Docker Repository
- Docker Username – Enter the username for the Docker Repository.
- Docker Password – Enter the password for the Docker Repository.



Important: Do not use the \$ character for this password.

- Docker Certificate – Click the Choose File button to upload a TLS certificate to secure communications with the Docker Repository.

Click the Generate the copy-docker script button to generate and download a script that copies the Docker images from Cloudera, or (for air-gapped installation) from a local http mirror in your network.

Run the script from a machine that is running Docker locally and has access to the Docker images using the following commands:

```
docker login [***URL for Docker Repository***] -u [***username of user with write access***]

bash copy-docker.txt
```

The copying operation may take 4 - 5 hours.

- 13.** In the Configure Data Services page, you can modify configuration settings such as the data storage directory, number of replicas, and so on. If there are multiple disks mounted on each host with different characteristics (HDD and SSD), then Local Path Storage Directory must point to the path belonging to the optimal storage. Ensure that you have reviewed your changes. If you want to specify a custom certificate, place the certificate and the private key in a specific location on the Cloudera Manager server host and specify the paths in the input boxes labelled as Ingress Controller TLS/SSL Server Certificate/Private Key File below. This certificate will be copied to the Control Plane during the installation process.

Click Continue.

Add Private Cloud Containerized Cluster

Configure Data Services

The Private Cloud Containerized Cluster needs to act as a TLS/SSL Server. By default, Cloudera Manager generates a self-signed certificate and uses it for all communication for example from the browser to the Private Cloud Containerized Cluster using TLS. If you want to specify a custom certificate, place the certificate and the private key in a specific location on the Cloudera Manager server host and specify the paths in the input boxes labelled as Ingress Controller TLS/SSL Server Certificate/Private Key File, below. This certificate must be valid for the application domain and one level underneath it. For example, if your application domain is 'apps.example.com', you must provide a wildcard certificate '*apps.example.com'. The certificate will be copied to the Private Cloud Containerized Cluster during the installation process.

Data Storage Directory	DOCKER (Service-Wide)
defaultDataPath	/docker
Edit Individual Values	
defaultDataPath	
	ECS (Service-Wide)
	/ecs/longhorn-storage
Application Domain	ECS (Service-Wide)
app_domain	
app_domain	
Local Path Storage Directory	ECS (Service-Wide)
isoDataPath	/ecs/local-storage
isoDataPath	
NFS Reserved Space	ECS (Service-Wide)
nfs_provisioned	<input type="text"/> GB
Number of Replicas	ECS (Service-Wide)
longhorn_replication	2
longhorn_replication	
Cluster IP Range	ECS (Service-Wide)
cluster_cidr	10.42.0.0/16
cluster_cidr	
Service IP Range	ECS (Service-Wide)
service_cidr	10.43.0.0/16
service_cidr	
Ingress Controller TLS/SSL Server Certificate File (PEM Format)	ECS (Service-Wide)
ssl_certificate	
ssl_certificate	
Ingress Controller TLS/SSL Server Private Key File (PEM Format)	ECS (Service-Wide)
ssl_private_key	
ssl_private_key	

- 14.** In the Configure Databases page, follow the instructions in the wizard to use your external existing databases with CDP Private Cloud.

Click Continue.

Add Private Cloud Containerized Cluster

- Getting Started
- Cluster Basics
- Specify Hosts
- Select JDK
- Enter Login Credentials
- Install Agents
- Assign Roles
- Configure Docker Repository
- Configure Data Services
- Configure Databases**
- Install Parcels
- Inspect Cluster
- Install Data Services
- Summary

Configure Databases

CDP Private Cloud Data Services uses databases for environments and apps metadata. You can connect to existing databases or create new databases with this wizard. [Learn more](#) about database requirements in CDP Private Cloud Data Services. If you choose the 'Use existing databases' option, the existing database server must be a PostgreSQL database server running version 10.6 or higher.

☒ Create embedded databases
☐ Use existing databases (Recommended for production)

Embedded Database Disk Space (GiB)

10

For production environments, Cloudera recommends that you use databases that you have previously created. These databases must all be on the same host and that host must be a PostgreSQL database server running version 10 or 12.

Ensure that you have selected the Use TLS for Connections Between the Control Plane and the Database option if you have plans to use Cloudera Data Warehouse (CDW). Enabling the Private Cloud Base Cluster PostgreSQL database to use an SSL connection to encrypt client-server communication is a requirement for CDW in CDP Private Cloud.

- 15.** In the Install Parcels page, the selected parcels are downloaded and installed on the host cluster. Click Continue.

Add Private Cloud Containerized Cluster

- Getting Started
- Cluster Basics
- Specify Hosts
- Select JDK
- Enter Login Credentials
- Install Agents
- Assign Roles
- Configure Docker Repository
- Configure Data Services
- Configure Databases
- Install Parcels**
- Inspect Cluster
- Install Data Services
- Summary

Install Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

> Embedded Container Service 1.4.0

Downloaded: 100%	Distributed: 3/3 (82.4 MB/s)	Unpacked: 3/3	Activated: 3/3
------------------	------------------------------	---------------	----------------

16. In the Inspect Cluster page, you can inspect your network performance and hosts. If the inspect tool displays any issues, you can fix those issues and run the inspect tool again.

Click Continue.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary

Inspect Cluster

You have created a new empty cluster. Cludera recommends that you run the following inspections. For accurate measurements, Cludera recommends that they are performed sequentially.

Inspect Network Performance

Advanced Options

You can use this tool to evaluate the network performance between hosts, such as ping latency.

Ping Timeout

10

Seconds

Ping Count

10

Ping Packet Size

56

Bytes

Status

Last Run

a few seconds ago

Duration

5.48s

Show Inspector Results

Run Again

More

Inspect Hosts

Once the inspection is complete, review the inspector results before proceeding.

Completed 0 of 3 step(s).

☐ I understand the risks of not running the inspections or the detected issues, let me continue with cluster setup.

17. In the Install Data Services page, you will see the installation process.

Add Private Cloud Containerized Cluster

Getting Started

Cluster Basics

Specify Hosts

Select JDK

Enter Login Credentials

Install Agents

Assign Roles

Configure Docker Repository

Configure Data Services

Configure Databases

Install Parcels

Inspect Cluster

Install Data Services

Summary

Install Data Services

First Run Command

Status

Running

Context

Containerized Cluster 1

May 9, 7:41:45 AM

Abort

Completed 0 of 1 step(s).

Show All Steps

Show Only Failed Steps

Show Only Running Steps

Run a set of services for the first time.

8/1 steps completed.

Execute 2 steps in sequence

8/1 steps completed.

Start DOCKER

8/1 steps completed.

Execute 3 steps in sequence

Waiting for command (Copy Images to Docker Registry (1546335347)) to finish

Execute 2 steps in sequence

Successfully executed command Generate Docker Certificate on service DOCKER

Execute command Prepare Host on service DOCKER

DOCKER

May 9, 7:41:45 AM

0ms

Execute command Generate Docker Certificate on service DOCKER

DOCKER

May 9, 7:41:45 AM

11.13s

Start DOCKER

Successfully started service.

DOCKER

May 9, 7:41:56 AM

22.45s

Starting 3 roles on service

May 9, 7:41:56 AM

22.45s

Execute 2 steps in sequence

Waiting for command (Copy Images to Docker Registry (1546335347)) to finish

Execute command Bring up Docker Registry on service DOCKER

DOCKER

May 9, 7:42:19 AM

3.44s

Execute command Copy Images to Docker Registry on service DOCKER

DOCKER

May 9, 7:42:22 AM

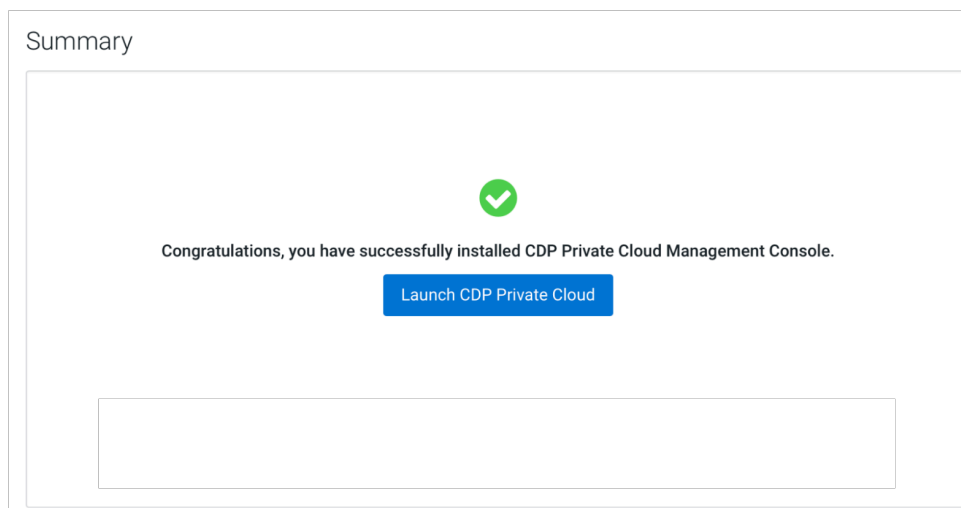
Abort

Start ECS

Execute 2 steps in sequence

27

18. After the installation is complete, you will see the Summary image. You can Launch CDP Private Cloud.



19. After the installation is complete, you can access your Private Cloud Data Services instance from Cloudera Manager > click Open Private Cloud Data Services.

If the installation fails, and you see the following error message in the stderr output during the Install Longhorn UI step, retry the installation by clicking the Resume button.

```
++ openssl passwd -stdin -apr1 + echo 'cm-longhorn:$apr1$gp2nrbtq$1KYPGI0QN1
FJ2lo5sV62l0' + kubectl -n longhorn-system create secret generic basic-auth
--from-file=auth + rm -f auth + kubectl -n longhorn-system apply -f /opt/clo
udera/cm-agent/service/ecs/longhorn-ingress.yaml Error from server (Internal
Error): error when creating "/opt/cloudera/cm-agent/service/ecs/longhorn-ing
ress.yaml":
Internal error occurred: failed calling webhook "validate.nginx.ingress.kub
ernetes.io": Post "https://rke2-ingress-nginx-controller-admission.kube-syst
em.svc:443/networking/v1/ingresses?timeout=10s": x509: certificate signed by
unknown authority
```

What to do next

- Click Open Private Cloud Data Services to launch your Private Cloud Experiences instance.
- Log in using the default username and password admin.
- In the Welcome to CDP Private Cloud page, click Change Password to change the Local Administrator Account password.
- Set up external authentication using the URL of the LDAP server and a CA certificate of your secure LDAP. Follow the instructions on the Welcome to CDP Private Cloud page to complete this step.
- Click Test Connection to ensure that you are able to connect to the configured LDAP server.
- [Create your first Virtual Warehouse in the CDW Data Service](#)
- [Provision an ML Workspace in the CML Data Service](#)
- [Add a CDE service in the CDE Data Service](#)

ECS Server High Availability

If you want to enable ECS Server for High Availability after installing ECS, then you must proceed with this section. If you do not want to enable ECS HA, you can safely ignore this section. You must review the note section and understand the ECS Server scenarios that are supported before you proceed to the next section.



Note:

- Longhorn replication defaults to two replicas. This can be set only during the installation time. Three or more replicas potentially have performance issues.
- Kubectl delete node <host> permanently removes host from cluster and any data on the host is lost. You must reformat the host before rejoining to the cluster.
- Single node failure may cause the Control Plane or any other management service to be unavailable. In 1.3.4 or later, it will take several minutes to recover automatically.

ECS Server scenarios

Clusters with only two servers are not supported. This is only for the temporary transition from a single server cluster to a three server cluster.

1. Three or more servers

- Redundancy requirements:
 - One failure requires three or more servers
 - Two failures require five or more servers
 - For more information see, [Fault Tolerance](#)
- To recover, you must scale-up the ECS Server roles. For more information on adding ECS node to a cluster, see the following section.

2. Two servers to one server

- Only after a double failure in a three server cluster
- To recover:
 - Stop the ECS service
 - Remove both the failed ECS server roles and hosts from cluster
 - On the surviving server, run the following command `/opt/cloudera/parcels/ECS/bin/rke2 server --cluster-reset`
 - Start the ECS service

3. Single server

- No failure supported

Enable ECS Server HA Post Installation

If you want to enable ECS Server for High Availability after installing ECS, then you must proceed with this section. If you do not want to enable ECS HA, you can safely ignore this section.

As a prerequisite, during the installation, you must have installed ECS with 1 master (with app_domain as Load Balancer URL) + agents. When you are adding more masters, ensure that you add Docker server as well.

Adding hosts to the containerised cluster

You must add hosts to the containerised cluster.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Click the Actions drop-down.
4. Click the Add Hosts button. The Add Hosts page appears.
5. Select the Add hosts to cluster option.
6. Select the cluster where you want to add the host from the drop-down list. Click Continue.

7. In the Specify Hosts page, provide a list of available hosts or you can add new hosts. You can provide the Fully Qualified Domain Name (FQDN) in the following patterns: You can specify multiple addresses and address ranges by separating them by commas, semicolons, tabs, or blank spaces, or by placing them on separate lines. Use this technique to make more specific searches instead of searching overly wide ranges.

For example, use host[1-3].network.com to specify these hosts: host1.network.com, host2.network.com, host3.network.com.

Click Continue.

8. In the Select Repository page, you must specify the repository location. Choose any one of the following:
 - a. Cloudera Repository (Requires direct internet access on all hosts)
 - b. Custom Repository
9. In the Select JDK page, select any one from the below options:
 - a. Manually manage JDK
 - b. Install a Cloudera-provided version of OpenJDK
 - c. Install a system-provided version of OpenJDK
10. In the Enter Login Credentials page select the SSH Username and provide the password.
11. The Install Agents page appears. Click Continue.
12. In the Install Parcels page, the selected parcels are downloaded and installed on the host cluster. Click Continue.
13. In the Inspect Hosts page, you can inspect your hosts. If the inspect tool displays any issues, you can fix those issues and run the inspect tool again. Click Continue.
14. In the Select Host Template page, select the hosts.
15. The Deploy Client Config page appears. Click Finish.

Adding Role Instances to Docker Server

You must add role instances to the docker server.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Open Docker Server.
4. Click the Actions drop-down.
5. Click the Add Role Instances button.
6. Select the hosts.
7. Click OK.

Adding Role Instances to Containerised Cluster

You must add the role instances to the containerised cluster.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Click the Actions drop-down.
4. Click the Add Role Instances button. The Add Role Instances page appears.
5. In the Assign Roles page, specify the role assignments for your new roles. Click Continue.
6. In the Review Changes page, click Finish.

Starting Docker Server on Nodes

You must start the Docker server on nodes.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Open Docker Server.
4. Click the Actions for Selected drop-down.
5. Click Start. Docker Server starts.

Starting ECS Server on Nodes

You must start the ECS server on nodes.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Click the Instances tab.
4. Select the nodes by clicking the checkbox
5. Click the Actions for Selected drop-down.
6. Click Start. ECS Server starts.

Refreshing ECS

You must refresh the ECS servers.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Click the Actions drop-down.
4. Click the Refresh button.

Checking Nodes and Pods in the UI

You must check the nodes and pods in the UI.

1. Log in to Cloudera Manager.
2. Navigate to the ECS service.
3. Click the Web UI drop-down.
4. Click ECS Web UI. The Kubernetes web UI page opens in a new tab.
5. Check the Nodes and Pods on the Web UI.

Enable ECS Server HA and promote agents Post ECS Installation

If you want to enable ECS Server for High Availability after installing ECS, then you must proceed with this section. If you do not want to enable ECS HA, you can safely ignore this section.

As a prerequisite, during the installation, you must have installed ECS with 1 master (with app_domain as Load Balancer URL) + agents. This allows you to promote Agents as masters.

Enabling ECS Server deployment for High Availability

Learn how to enable ECS Server deployment for High Availability by installing a Load Balancer and promoting the existing ECS Agents to ECS Server. By performing this procedure, you will be able to deploy HA on your existing ECS Server.

If you have a production quality ECS cluster, then Cloudera recommends you to use ECS Server High Availability. You can also consider having an ECS Server HA for any non-production ECS cluster that you expect to be available long-term.

If you have Cloudera Manager 7.5.5, then you must have an ECS cluster installed and configured with a single ECS Server.

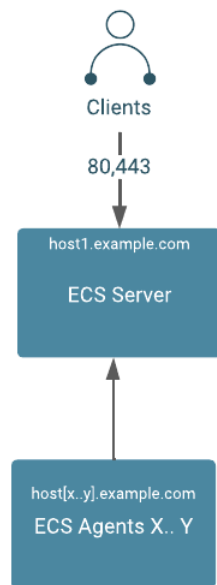
You can enable an ECS cluster and convert it to a cluster protected by ECS Server HA. Enabling ECS Server deployment for High Availability involves preparing your cluster, configuring DNS wildcard entry, adding a Load Balancer into the topology, and promoting ECS Agents to the ECS Server. An ECS High Availability cluster must consist of:

- An odd number of server nodes that will run etcd, the Kubernetes API, and other control plane services. Cloudera recommends a minimum of three ECS Server nodes.
- Two or more agent nodes that are designated to run CDP data services.
- A software or hardware Load balancer using TCP mode (non-terminating https).

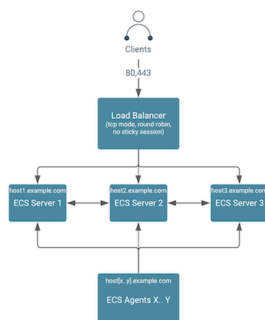


Note: A Load Balancer is required for the ECS Server HA. This documentation uses HAProxy as an example. However, Cloudera recommends that you use your production quality Load Balancer technology from commercial vendors.

Architecture of CDP Private Cloud Experiences on single ECS Server



Architecture of CDP Private Cloud Experiences on High Availability



Preparing the cluster for High Availability:

Review the table to understand the requirements for enabling the High Availability.

1. This process has been tested with a minimum of five ECS hosts. However, Cloudera recommends six or more hosts.
2. DNS requirements for ECS High Availability must be fulfilled.

Hostname	Subdomain	Expected Roles	DNS ForwardZone	Reverse Zone PTR
“Wildcard” (hostname = *)	apps.ecs.example.com The string “apps” is required, “ecs” is up to user	Virtual app domain wildcard	“A Record” wildcard (hostname = *), may be a CNAME on certain DNS systems that use text-based config. Resolves to fixed IP of ha_proxy (or VIP of some commercial LB’s)	N
“apps alias”	apps.ecs.example.com	Virtual app domain alias	“CNAME” alias points to A Record of ha_proxy (or VIP). Alternatively, this can be an ARecord with IP of ha_proxy (or VIP)	N/A
HAProxy (or commercial LB)	<domain of your LB>	HA Load Balancer	Depends on vendor/ software	

Hostname	Subdomain	Expected Roles	DNS ForwardZone	Reverse Zone PTR
ecs-master1	example.com	ECS Server 1 Docker server	"A Record" resolves to IP of ecs-master1	Y
ecs-master2	example.com	ECS Server 2 Docker server	"A Record" resolves to IP of ecs-master2	Y
ecs-master3	example.com	ECS Server 3 Docker server	"ARecord" resolves to IP of ecs-master3	Y
ecs-agentN	example.com	ECS Agent N Docker server N	"ARecord" resolves to IP of ecs-agentN	Y

**Note:**

1. The above table uses a consistent subdomain ("example.com") but this is not mandatory. To support multiple domains, you must follow certain steps to ensure that the domains are forward and reverse resolvable using DNS, from all Base cluster and ECS cluster hosts (that is through forest/domain level trusts and/or hosts level /etc/resolv.conf config). You must avoid the use of /etc/hosts entries.
2. A predefined wildcard DNS record allows the resolution of *.apps.<app domain name> to the IP address of the Load Balancer. You cannot proceed further until this is in place.

High Level steps to enable an ECS High Availability cluster

Review the high level steps to understand the steps in enabling High Availability.

Enabling ECS High Availability Cluster

- 1 [Verifying DNS Setup](#)
- 2 [Installing Load Balancer](#)
- 3 [Promoting ECS Agents to ECS Servers](#)
- 4 [Refreshing ECS Cluster](#)

**Note:**

1. You must have installed an ECS with one ECS server and other nodes that are ECS Agents.
2. You must have a DNS wildcard record that has an IP address pointing to your Load Balancer (hostname or VIP). For more information, see the [KB article](#).

Verifying DNS setup

You must verify the DNS setup to ensure that the app domain DNS hostname points to the Load Balancer.

Procedure

1. Verify that the app domain DNS hostname has moved from single non-HA ECS Server to the Load Balancer.

Hostname	Expected Roles	DNS
ecs-loadbalancer.example.com	Load Balancer	Resolves to IP of LB host (or VIP). The example uses 10.10.0.99. Both *.apps.ecs.example.com and apps.ecs.example.com resolve to 10.10.0.99.

2. Verify the DNS setup with nslookup.



Note: You must verify that a random hostname resolves in the wildcard entry. In this example, Cloudera uses foobar.apps.ecs.example.com as the random name. Both entries should resolve to the same IP address.

For example,

```
$ hosts="apps.ecs.example.com foobar.apps.ecs.example.com"
$ for target in $hosts; do nslookup $target; done

Server: 10.10.xx.xx
Address: 10.10.xx.xx#53

apps.ecs.example.com canonical name = ecs-loadbalancer.example.com.
Name: ecs-loadbalancer.example.com
Address: 10.10.0.99

Server: 10.10.xx.xx
Address: 10.10.xx.xx#53

Name: foobar.apps.ecs.example.com
Address: 10.10.0.99
```

Results

DNS setup is verified.

What to do next

You must now install the Load Balancer.

Installing Load Balancer

To install the HAProxy Load Balancer, Cloudera uses an example that uses a single instance of HAProxy, configured with round robin balancing and TCP mode. This allows for non-terminating https (https passthrough). The HAProxy service can be configured for High Availability using keepalived.

Before you begin

You must consult your operating system vendor's documentation for requirements and the install guide for configuring HAProxy with keepalived.

To install a HAProxy Load Balancer, you must ssh into the HAProxy host, install, and then configure HAProxy:

Procedure

1. `sudo su -`
2. `yum install haproxy -y`
3. `cp /etc/haproxy/haproxy.cfg /etc/haproxy/haproxy.cfg.bak`

4. `cat > /etc/haproxy/haproxy.cfg << EOF`

global

log	127.0.0.1 local2
chroot	/var/lib/haproxy
pidfile	/var/run/haproxy.pid
user	haproxy
group	haproxy
daemon	

defaults

mode	tcp
log	global
option	tcplog
option	dontlognull
option	redispatch
retries	3
maxconn	5000
timeout connect	5s
timeout client	50s
timeout server	50s

listen stats

bind *:8081
mode http
stats enable
stats refresh 30s
stats uri /stats
monitor-uri /healthz

frontend fe_k8s_80

bind *:80
default_backend be_k8s_80

backend be_k8s_80

balance roundrobin
mode tcp
server ecs-server1.example.com 10.10.0.1:80 check
server ecs-server2.example.com 10.10.0.2:80 check

```
server ecs-server3.example.com 10.10.0.3:80 check
```

```
frontend fe_k8s_443
```

```
bind *:443
```

```
default_backend be_k8s_443
```

```
backend be_k8s_443
```

```
balance roundrobin
```

```
mode tcp
```

```
server ecs-server1.example.com 10.10.0.1:443 check
```

```
server ecs-server2.example.com 10.10.0.2:443 check
```

```
server ecs-server3.example.com 10.10.0.3:443 check
```

EOF

```
systemctl enable haproxy
```

```
systemctl restart haproxy
```

```
systemctl status haproxy
```

5. You can verify that all the hosts are shown from the HAProxy UI. However, at this point the hosts are not listening to the configured ports.

Stats																																				
		Queue			Session rate			Sessions					Bytes			Denied			Errors			Warnings			Status			LastChk			Weight			Server		
	Frontend	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Weight	Act	Bck	Chk	Dwn	Downtime	Thrtle					
	Backend	0	0	0	0	0	0	-	1	1	5 000	0	0	0	0	0	0	0	0	0	0	0	OPEN		0	0	0	0	0	0	0	0	0			
fe_k8s_80																																				
		Queue			Session rate			Sessions					Bytes			Denied			Errors			Warnings			Status			LastChk			Weight			Server		
	Frontend	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Weight	Act	Bck	Chk	Dwn	Downtime	Thrtle					
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	OPEN		0	0	0	0	0	0	0	0	0			
be_k8s_80																																				
		Queue			Session rate			Sessions					Bytes			Denied			Errors			Warnings			Status			LastChk			Weight			Server		
	Frontend	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Weight	Act	Bck	Chk	Dwn	Downtime	Thrtle					
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	Backend	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
fe_k8s_443																																				
		Queue			Session rate			Sessions					Bytes			Denied			Errors			Warnings			Status			LastChk			Weight			Server		
	Frontend	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Weight	Act	Bck	Chk	Dwn	Downtime	Thrtle					
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	OPEN		0	0	0	0	0	0	0	0	0			
be_k8s_443																																				
		Queue			Session rate			Sessions					Bytes			Denied			Errors			Warnings			Status			LastChk			Weight			Server		
	Frontend	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Weight	Act	Bck	Chk	Dwn	Downtime	Thrtle					
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
		0	0	0	0	0	0	-	0	0	5 000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	Backend	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			



Important: Since you already have an ECS cluster running, you must alter your DNS wildcard to point to the IP address of the HAProxy server. You cannot change the Application Domain configured through the ECS wizard. So you must ensure that you send all ingress traffic to the HAProxy IP address by making that change in the IP address of your wildcard DNS Record.



Note:

- Application Domain (app_domain property in Cloudera Manager) maps to your wildcard DNS record (For example, app_domain ecs.example.com maps to your DNS entry *.apps.ecs.example.com)
- The resolved IP address must be the host IP (or VIP) of your Load Balancer. For more information, see the Verify DNS Step 5 above.

Results

Load Balancer is now installed.

Promoting ECS Agents to ECS Servers

You must promote ECS Agents to ECS Servers.

About this task

After installing the Load Balancer, you must reconfigure the existing ECS Agents to ECS Servers. This process is called Promoting the agents to Servers. You must promote only one agent at a time. This procedure is explained using an example where you promote the ECS agent on agent1.example.com and then promote the ECS agent on agent2.example.com.

Procedure

- 1. Prepare the agent node for promotion by running the following commands on the command line of your ECS Server host.
sudo /var/lib/rancher/rke2/bin/kubect1 --kubeconfig=/etc/rancher/rke2/rke2.yaml get nodes
sudo /var/lib/rancher/rke2/bin/kubect1 --kubeconfig=/etc/rancher/rke2/rke2.yaml drain agent1.example.com --ignore-daemonsets --delete-emptydir-data.



Note: This may take a few minutes.

- 2. In Cloudera Manager, navigate to ECS Cluster >> ECS. Stop the ECSAgent running on agent1 and then delete the agent.

ECS-HACluster-01

ECS

Status

Instances

Configuration

Commands

Charts Library

Audits

Web UI

Quick Links

Q Search

Filters

Last Updated: Jan 30, 10:01:58 PM PST

Add Role Instances

Role Groups

Filters

STATUS

Good Health 5

Stopped 1

ROLE GROUP

ROLE TYPE

STATE

HEALTH TEST

Actions for Selected (1)

<input type="checkbox"/>	Status	Role Type	State	Hostname	Commission State	Role Group
<input type="checkbox"/>	✔	Ecs Agent	Started	[redacted].com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>	✔	Ecs Agent	Started	[redacted].com	Commissioned	Ecs Agent Default Group
<input checked="" type="checkbox"/>	⊖	Ecs Agent	Stopped	[redacted].com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>	✔	Ecs Agent	Started	[redacted].com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>	✔	Ecs Agent	Started	[redacted].com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>	✔	Ecs Server	Started	[redacted].com	Commissioned	Ecs Server Default Group

1 - 6 of 6

- 3. In Cloudera Manager, navigate to ECS Cluster >> ECS. Click Add Role Instances.

Add Role Instances to ECS

1 Assign Roles

2 Review Changes

Assign Roles

You can specify the role assignments for your new roles here.

You can also view the role assignments by host:

View By Host

Ecs Server × 1

Ecs Agent × 3

Select hosts

Select hosts

4. Add the available host agent1 as an ECS Server in the Add Role Instances to ECS pop-up. Click Ok.

Add Role Instances to ECS

3 Hosts Selected

Select hosts for a new or existing role. The host list is filtered to remove hosts that are not valid candidates; these include hosts that are unhealthy, members of other clusters, or have an incompatible version of the software installed on them.

Q Enter hostnames: host01, IP addresses or rack

<input type="checkbox"/>	Hostname	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
<input checked="" type="checkbox"/>	clev-ws1.sme-clev.athens.cloudera.com	10.113.207.11	/default	16	62.8 GIB	DS ES	ES
<input checked="" type="checkbox"/>	<input type="text"/> .com	10.113.207.12	/default	16	62.8 GIB	DS	ES
<input checked="" type="checkbox"/>	<input type="text"/> .com	10.113.207.13	/default	16	62.8 GIB	DS	ES
<input type="checkbox"/>	<input type="text"/> .com	10.113.207.14	/default	16	62.8 GIB	DS EA	
<input type="checkbox"/>	<input type="text"/> .com	10.113.207.18	/default	16	62.8 GIB	DS EA	
<input type="checkbox"/>	<input type="text"/> .com	10.113.207.19	/default	16	62.8 GIB	DS EA	

1 - 6 of 6

Cancel

OK

Add Role Instances to ECS

1 Assign Roles

2 Review Changes

Assign Roles

You can specify the role assignments for your new roles here.

You can also view the role assignments by host.

View By Host

Ecs Server × (1 + 2 New)

Ecs Agent × 3

Select hosts

5. Click Continue.

Experiences Cluster 1

ECS

Actions

Status

Instances

Configuration

Commands

Charts Library

Web UI

Quick Links

Q Search

Filters

Last Updated: Jan 6, 3:58:16 AM UTC

Add Role Instances

Role Groups

Filters

STATUS

Good Health4

Stopped2

ROLE GROUP

ROLE TYPE

STATE

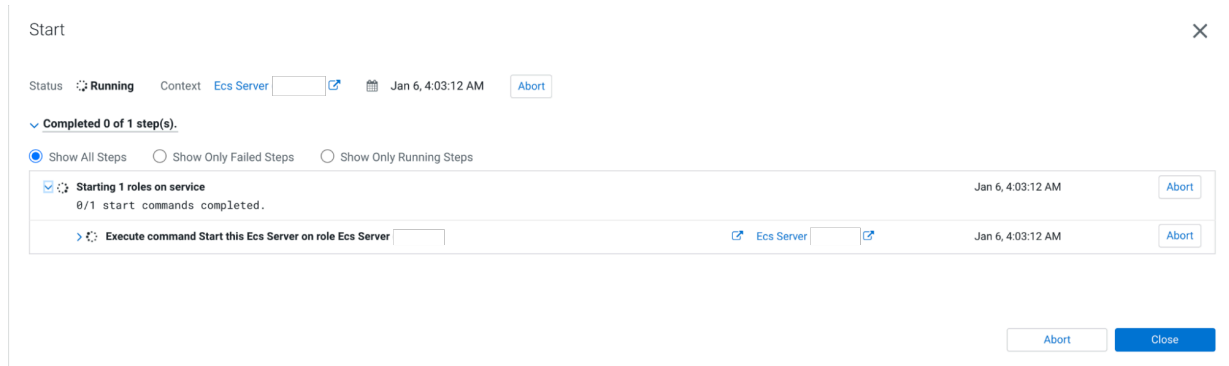
HEALTH TEST

Actions for Selected

<input type="checkbox"/>	Status	Role Type	State	Hostname	Commission State	Role Group
<input type="checkbox"/>		Ecs Agent	Started	<input type="text"/> .com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>		Ecs Agent	Started	<input type="text"/> .com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>		Ecs Agent	Started	<input type="text"/> .com	Commissioned	Ecs Agent Default Group
<input type="checkbox"/>		Ecs Server	Stopped	<input type="text"/> .com	Commissioned	Ecs Server Default Group
<input type="checkbox"/>		Ecs Server	Stopped	<input type="text"/> .com	Commissioned	Ecs Server Default Group
<input type="checkbox"/>		Ecs Server	Started	<input type="text"/> .com	Commissioned	Ecs Server Default Group

38

6. Start the new ECS Server from ECS Instances view. For example, start ECSServer on agent1.



7. On the command line, uncordon the node by running the following command: `sudo /var/lib/rancher/rke2/bin/kubect1 --kubeconfig=/etc/rancher/rke2/rke2.yaml uncordon agent1.example.com`
8. Confirm the node's status from webUI or the command line by running the command `sudo /var/lib/rancher/rke2/bin/kubect1 --kubeconfig=/etc/rancher/rke2/rke2.yaml get nodes`.



Note: Do not proceed until node status is Ready. This may take several minutes.

Name	Labels	Ready	CPU requests (cores)	CPU limits (cores)	Memory requests (bytes)	Memory limits (bytes)	Pods	Created
beta.kubernetes.io/arch: amd64	beta.kubernetes.io/os: linux ecs_role: master	True	4.54 (28.38%)	0.00m (0.00%)	0.00 (0.00%)	0.00 (0.00%)	12 (10.91%)	48 seconds ago

What to do next

When agent1 is ready, you can promote the next agent agent2. To promote the next agent, you must perform steps 1-8 again, the example uses agent2.example.com.

Refreshing ECS

After all the ECS Agents are promoted to ECS Servers, you must log in to Cloudera Manager and refresh the ECS cluster.

Procedure

- 1. Navigate to ECS Cluster >> ECS view >> Actions >> Refresh ECS. This sets the ingress proxy so that all three servers are eligible to process incoming commands.

Experiences Cluster 1

ECS

Actions

Start

Stop

Restart

Add Role Instances

Rename

Delete

Enter Maintenance Mode

Unseal Vault

Reapply All Settings to Cluster

Update Ingress Controller Certificate

Refresh ECS

Create Environment

Status

Instances

Health Tests

Show 7 Good

Status Summary

Ecs Agent

Ecs Server

Hosts

Health History

Ecs Server Health

Ecs Server Health Bad

Charts

Informational Events

Important Events and Alerts

Refresh ECS

Are you sure you want to refresh the ECS service?

Cancel Refresh ECS

Refresh ECS

Status Finished Context ECS Jan 7, 5:56:31 PM 3.69s

Successfully refreshed the ECS service.

Completed 4 of 4 step(s).

Show All Steps Show Only Failed Steps Show Only Running Steps

Execute command Refresh Ecs Server on role Ecs Server

Ecs Server

Jan 7, 5:56:31 PM

15ms

Execute command Refresh Ecs Server on role Ecs Server

Ecs Server

Jan 7, 5:56:31 PM

3ms

Execute command Refresh Ecs Server on role Ecs Server

Ecs Server

Jan 7, 5:56:31 PM

3ms

Execute command Reapply All Settings to Cluster on service ECS

ECS

Jan 7, 5:56:31 PM

3.63s

Close

2. Confirm that all backends of HAProxy display the status UP. This may take several minutes.

#5555																								
Queue					Session rate					Sessions					Bytes					Denied				
Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	Limit	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Wght
Frontend	1	2	-	1	2	5 000	144						132 493	3 570 185	0	0	0	0	0	0	0	OPEN		0
Backend	0	0		0	1	0	1	5 000	143	0	0s		132 493	3 570 185	0	0	0	0	0	0	0	1h12m UP		0

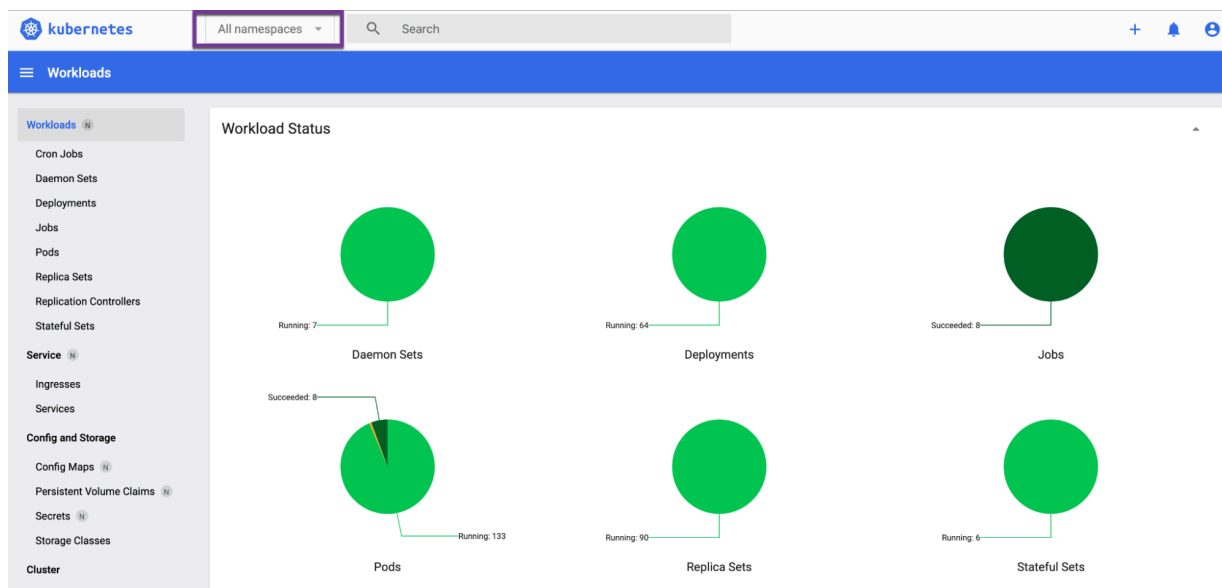
#5555_80																								
Queue					Session rate					Sessions					Bytes					Denied				
Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	Limit	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Wght
Frontend	0	0	-	0	0	5 000	144						0	0	0	0	0	0	0	0	0	OPEN		0
Backend	0	0	-	0	0	0	0	5 000	143	0	0s		0	0	0	0	0	0	0	0	0	1h12m UP		0

#5555_80																								
Queue					Session rate					Sessions					Bytes					Denied				
Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	Limit	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Wght
com	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32m23s UP	L4OK in 0ms	1
com	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15m44s UP	L4OK in 0ms	1
com	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15m44s UP	L4OK in 0ms	1
Backend	0	0	-	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	32m23s UP		3

#5555_443																								
Queue					Session rate					Sessions					Bytes					Denied				
Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	Limit	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Wght
Frontend	0	24	-	3	8	5 000	493						901 947	2 478 032	0	0	0	0	0	0	0	OPEN		0
Backend	0	0	-	0	0	0	0	5 000	493	0	0s		901 947	2 478 032	0	0	0	0	0	0	0	1h12m UP		0

#5555_443																								
Queue					Session rate					Sessions					Bytes					Denied				
Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	Limit	LiTot	Last	In	Out	Req	Resp	Req	Conn	Resp	Retr	Redis	Status	LastChk	Wght
com	0	0	-	0	8	1	4	1	4	1	261	261	47s	430 509	1 502 801	0	0	0	0	0	0	32m24s UP	L4OK in 0ms	1
com	0	0	-	0	8	1	3	1	3	1	114	114	42s	233 867	478 225	0	0	0	0	0	0	15m43s UP	L4OK in 0ms	1
com	0	0	-	0	8	1	3	1	3	1	114	114	42s	237 571	497 006	0	0	0	0	0	0	15m45s UP	L4OK in 0ms	1
Backend	0	0	-	0	24	3	8	500	493	489	42s	901 947	2 478 032	0	0	0	0	0	0	0	0	32m24s UP		3

3. Confirm that all pods are green in the ECS webUI >> (All Namespaces) >> Workloads.



4. Confirm that there are no alerts in the ECS service.

ECS1

The screenshot displays the ECS1 dashboard. At the top, there is a header with a green checkmark icon, a gear icon, the text 'ECS', and an 'Actions' dropdown menu. Below this is a navigation bar with tabs: 'Status' (underlined), 'Instances', 'Configuration', 'Commands' (with a blue play button icon and the number '1'), and 'Charts Library'. The main content area is titled 'Health Tests' and includes a 'Create Trigger' button. Below the title, a box shows a green checkmark icon and the text 'Show 7 Good'. Further down, the 'Status Summary' section contains a table with three rows: 'Ecs Agent' with '1 Good Health', 'Ecs Server' with '3 Good Health', and 'Hosts' with '4 Good Health'.

Component	Health Status
Ecs Agent	1 Good Health
Ecs Server	3 Good Health
Hosts	4 Good Health

Results

High Availability is now deployed on your ECS cluster.

Manually uninstalling ECS from a cluster

You can manually uninstall ECS from your cluster.

Before you begin

Before performing this procedure, ensure that you have activated the ECS parcel on the cluster hosts.

During the installation time of ECS, the directory for Longhorn and the LSO are decided by Cloudera Manager and defaults to /ecs.

Data Storage Directory defaultDataPath Edit Individual Values defaultDataPath	DOCKER (Service-Wide) ⓘ <input type="text" value="/docker"/>
	ECS (Service-Wide) ⓘ <input type="text" value="/ecs/longhorn-storage"/>
Application Domain app_domain app_domain	ECS (Service-Wide) ⓘ <input type="text" value="cloudera.com"/>
Local Path Storage Directory IsoDataPath IsoDataPath	ECS (Service-Wide) ⓘ <input type="text" value="/ecs/local-storage"/>

Procedure

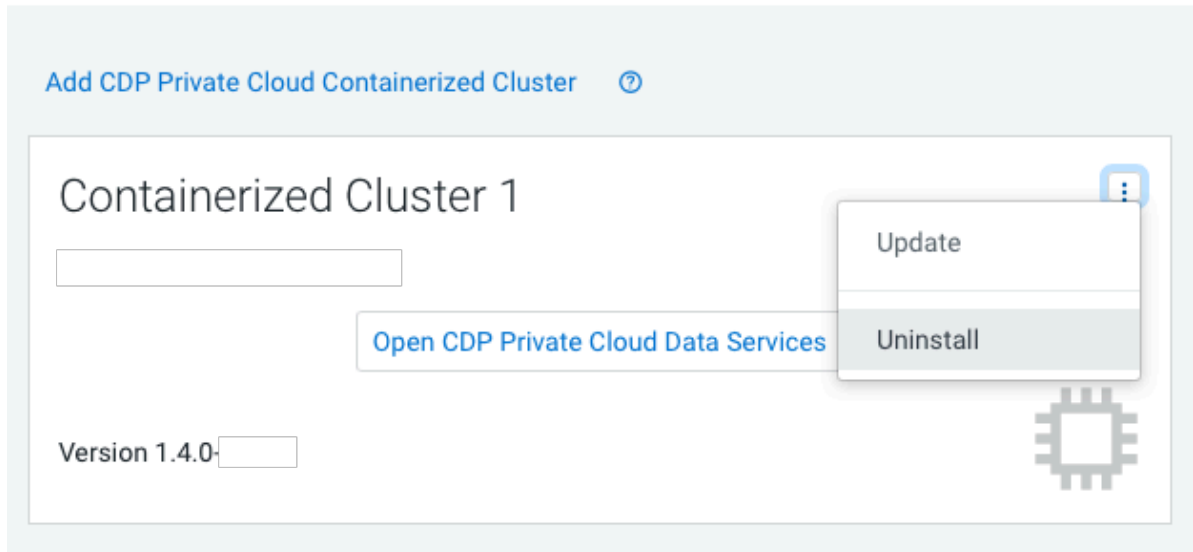
1. On each host in the cluster:
 - a) `/opt/cloudera/parcels/ECS/docker/docker container stop registry`
 - b) `/opt/cloudera/parcels/ECS/docker/docker container rm -v registry`
 - c) `/opt/cloudera/parcels/ECS/docker/docker image rm registry:2`
2. Stop the ECS cluster in Cloudera Manager
3. On each host:
 - a) `cd /opt/cloudera/parcels/ECS/bin`
 - b) `./rke2-killall.sh` # usually 2 times is sufficient
 - c) `./rke2-uninstall.sh`
 - d) `rm -rf /ecs/*` # assumes the default defaultDataPath and IsoDataPath
 - e) `rm -rf /var/lib/docker_server/*` # deletes the auth and certs
 - f) `rm -rf /etc/docker/certs.d/*` # delete the ca.crt
 - g) `rm -rf /docker` # assumes the default defaultDataPath for docker

4. Delete the ECS cluster in Cloudera Manager.

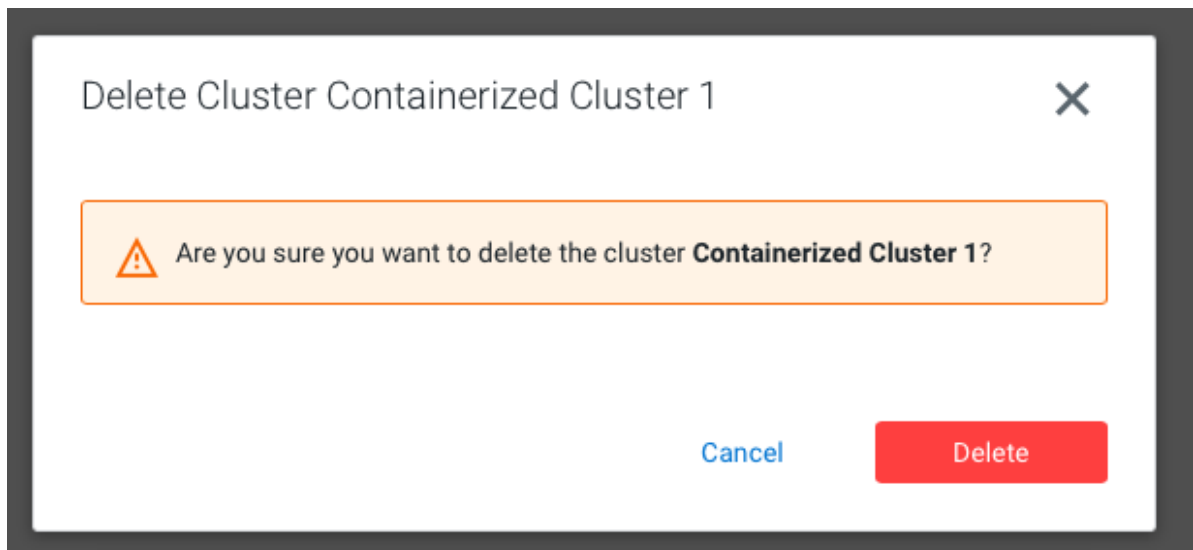
a)

In Cloudera Manager, navigate to CDP Private Cloud Data Services and click . Click Uninstall.

CDP Private Cloud Data Services



b) The Delete Cluster wizard appears. Click Delete.



5. Clean IPtables on each host:

```
echo "Reset iptables to ACCEPT all, then flush and delete all other chains";
declare -A chains=( [filter]=INPUT:FORWARD:OUTPUT
[raw]=PREROUTING:OUTPUT [mangle]=PREROUTING:INPUT:FORWARD:OUTPUT:POSTROUTING
[security]=INPUT:FORWARD:OUTPUT [nat]=PREROUTING:INPUT:OUTPUT:POSTROUTING );
for table in "${!chains[@]}"; do
echo "${chains[$table]}" | tr : $'\n' | while IFS=
do sudo iptables -t "$table" -P "$REPLY" ACCEPT
done
read -r;
```

```
sudo iptables -t "$table" -F
sudo iptables -t "$table" -X
done
```



Note: Alternatively, an experimental script is available. This script combines steps three through five. The script is available here: <https://github.com/cloudera-labs/snippets/blob/main/private-cloud/kill-2-rke.sh>

6. Reboot the host(s).
7. Before you install ECS again, ensure that the IP tables list is empty by executing the following command: `#iptables -F`

Upgrading

Upgrading Cloudera Manager

You must use the Cloudera Manager version 7.8.1 to set up the Private Cloud Experiences cluster.

If you already have a Private Cloud Base cluster setup using an earlier version of Cloudera Manager, you must first upgrade the Cloudera Manager version to Cloudera Manager 7.8.1 release and then begin the ECS installation.

For more information, see [Upgrading Cloudera Manager](#).

Related Information

[Upgrading Cloudera Manager](#)

Update from 1.3.4 or 1.4.0-H1 to 1.4.1 (ECS)

You can update your existing CDP Private Cloud Data Services 1.3.4 or 1.4.0-H1 to 1.4.1 without requiring an uninstall.

Before you begin

Upgrading the Embedded Container Service (ECS) version while CDE service is enabled, can cause Control Pane upgrade looping forever in error state. You must back up CDE jobs in the CDE virtual cluster, and then delete the CDE service and CDE virtual cluster. Restore it after the upgrade. For more information about backup and restore CDE jobs, see [Backing up and restoring CDE jobs](#).

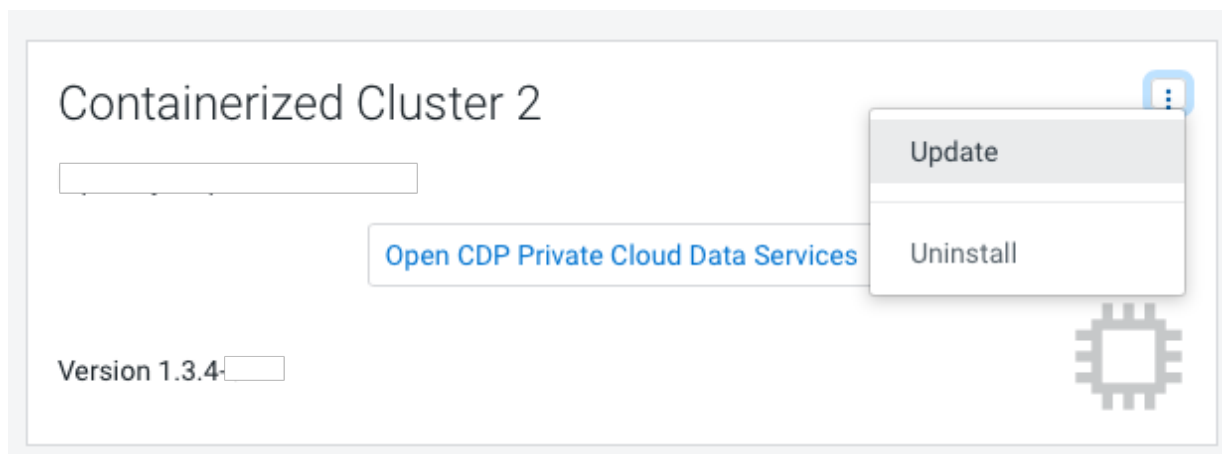
About this task



Note: This procedure requires down time during the upgrade.

Procedure**1.**

In Cloudera Manager, navigate to CDP Private Cloud Data Services and click . Click Update.



2. On the Getting Started page, you can select the Install method - Air Gapped or Internet and proceed.

Internet install method

Update Private Cloud Data Services (cdp)

1 Getting Started

2 Collect Information

3 Install Parcels

4 Update Data Services

5 Summary

Getting Started

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services.

Visit the [CDP Private Cloud](#) documentation for more information.

Current Version
1.3.4

Install Method
☒ Internet
 ☐ Air Gapped

1. Select Repository

Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater

You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version** update. Please make sure you have backed up all the external databases.

Air Gapped install method

Update Private Cloud Data Services (cdp)

1 Getting Started

2 Collect Information

3 Install Parcels

4 Update Data Services

5 Summary

Getting Started

This wizard provides step-by-step guidance for updating CDP Private Cloud Data Services.

Visit the [CDP Private Cloud](#) documentation for more information.

Current Version
1.3.4

Install Method
☐ Internet
 ☒ Air Gapped

Installing via a local mirror with an http server. You will need to setup a full mirror of Cloudera's repositories via a temporary http server within the perimeter network of all hosts.

- Download everything under `https://archive.cloudera.com/p/cdp-pvc-ds/latest`
- Modify the file `manifest.json` inside the downloaded directory, change `"http_url": "..."` to `"http_url": "http://your_local_repo/cdp-pvc-ds/latest"`
- Mirror the downloaded directory to your local http server, e.g. `http://your_local_repo/cdp-pvc-ds/latest`
- Add `http://your_local_repo/cdp-pvc-ds/latest` to your [Custom Repository](#) settings and select it from the dropdown below.
- Select Repository

Please ensure all the Data Lake clusters are running Cloudera Runtime 7.1.6 or greater

You are about to update CDP Private Cloud Data Services to version 1.4.0. This is a **minor version** update. Please make sure you have backed up all the external databases.

Click Continue.

3. On the Collect Information page, click Continue.

Update Private Cloud Data Services (cdp)

✓ Getting Started

2 Collect Information

3 Install Parcels

4 Update Data Services

5 Summary

Collect Information

Sometimes, new configuration information might be needed before you can update. If there are no configuration needed below, just click Next.

Configure Vault

Vault is a secret management tool. You can connect to an existing customer Vault or create a new Vault with this installer. [Learn more](#) on Vault on CDP Private Cloud Data Services.

☒ Embedded vault
 ☐ External Vault (Recommended for production)

4. On the Install Parcels page, click Continue.

Update Private Cloud Data Services (cdp)

Getting Started

Collect Information

Install Parcels

Update Data Services

Summary

Install Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

Embedded Container Service 1.4.0

All (1)

Running (1)

Failed (0)

Completed (0)

Downloaded: 100%

Distributed: 1/1 (3.9 MB/s)

Unpacked: 0/1

Hostname	Throughput	Status	Errors
kprranay-4.vpc.cloudera.com	9.9 MB/s	DISTRIBUTING	

48

5. On the Update Progress page, you can see the progress of your update. Click Continue after the update is complete .

Update Private Cloud Data Services (cdp)

Getting Started

Collect Information

Install Parcels

Update Data Services

Summary

Update Data Services

Upgrade Cluster Command

Status Running Context Containerized Cluster 2 May 9, 8:46:13 AM Abort

Completed 5 of 6 step(s).

Show All Steps

Show Only Failed Steps

Show Only Running Steps

Execute command Step on service ECS-2

Execute command Step on service DOCKER-2

Activating parcel

Waiting for Cloudera Manager Agents to detect release ECS 1.4.0.

Converting configuration parameters

Starting all services in the upgraded cluster.

ECS-2

DOCKER-2

Containerized Cluster 2

Containerized Cluster 2

Containerized Cluster 2

Containerized Cluster 2

Containerized Cluster 2

May 9, 8:46:13 AM

May 9, 8:46:14 AM

May 9, 8:46:16 AM

May 9, 8:46:16 AM

May 9, 8:46:31 AM

May 9, 8:46:31 AM

309ms

2.31s

68ms

15.07s

24ms

49



Note: The upgrade might occasionally fail with error messages or conditions such as the following:

- Error message: During the following step: Execute command Install Tolerations Webhook on service ECS-3 the Upgrade progress page mentions a failure waiting for kube-proxy to come up.

Workaround:

- a. Log in using ssh to one of the ECS Server nodes and run the following command:

```
/var/lib/rancher/rke2/bin/kubectl get nodes
```

The output will look similar to the following:

NAME	STATUS	ROLES
ecs-abc-1.vpc.myco.com	Ready	control-plane,etcd,master
4h50m v1.21.8+rke2r2		
ecs-abc-2.vpc.myco.com	NotReady	<none>
4h48m v1.20.8+rke2r1		
ecs-abc-3.vpc.myco.com	Ready	<none>
4h48m v1.21.8+rke2r2		
ecs-abc-4.vpc.myco.com	NotReady	<none>
4h48m v1.20.8+rke2r1		
ecs-abc-5.vpc.myco.com	NotReady	<none>
4h48m v1.20.8+rke2r1		

If any of the version numbers in the last column are lower than the expected version, reboot those nodes. (For example, v1.20.8 in the output above.)

- b. In the Command Output window, in the step that failed, click Resume.
- Upgrade hangs on the Execute command Post upgrade configuration on service ECS step for more than an hour.

Workaround:

- a. Log in to one of the ECS server nodes and run the following command:

```
kubectl get nodes
```

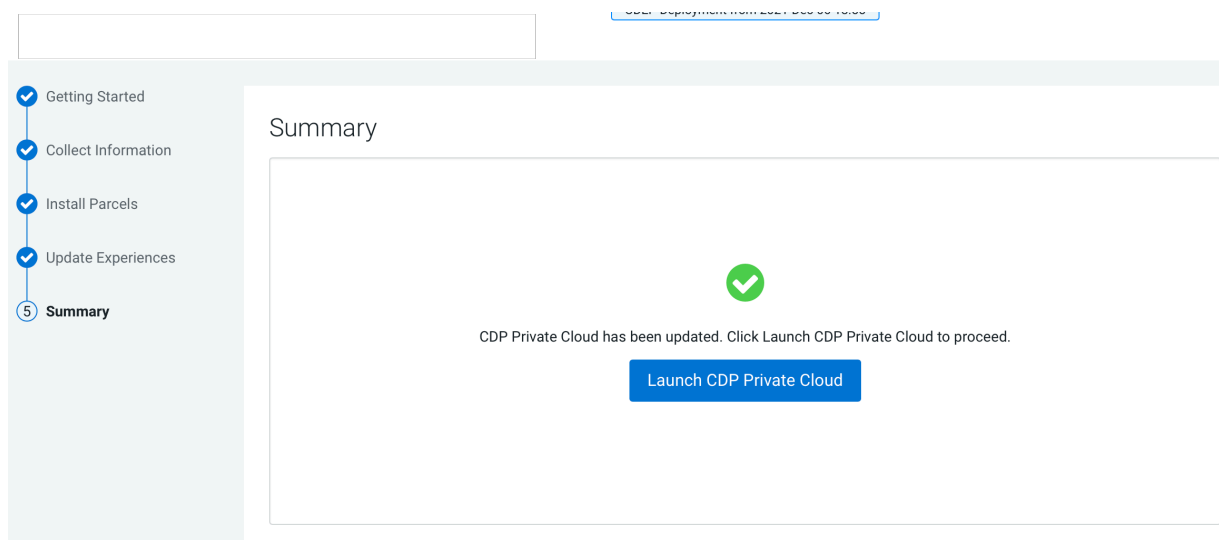
The output looks similar to the following:

NAME	STATUS	ROLES
ecs-abc-1.vpc.myco.com	Ready	control-plane,etcd,master
3h47m v1.21.11+rke2r1		
ecs-abc-2.vpc.myco.com	NotReady	<none>
3h45m v1.21.8+rke2r2		
ecs-abc-3.vpc.myco.com	NotReady	<none>
3h45m v1.21.8+rke2r2		
ecs-abc-4.vpc.myco.com	NotReady	<none>
3h45m v1.21.8+rke2r2		

If you any nodes display a status of NotReady, click the Abort button in the command output window.

- b. Reboot all nodes showing NotReady.
- c. Check the node status again as shown above. After all the nodes show Ready, click the Resume button in the command output window to continue with the upgrade.

6. After the update is complete, the Summary page appears. You can now Launch CDP Private Cloud from here.



If you see a Longhorn Health Test message about a degraded Longhorn volume, wait for the cluster repair to complete.

Or you can navigate to the CDP Private Cloud Data Services page and click Open CDP Private Cloud Data Services.

CDP Private Cloud Data Services opens up in a new window.

- If the upgrade stalls, do the following:

1. Check the status of all pods by running the following command on the ECS server node:

```
kubectl get pods --all-namespaces
```

2. If there are any pods stuck in “Terminating” state, then force terminate the pod using the following command:

```
kubectl delete pods <NAME OF THE POD> -n <NAMESPACE> --grace-period=0 -f  
orce
```

If the upgrade still does not resume, continue with the remaining steps.

3. In the Cloudera Manager Admin Console, go to the ECS service and click Web UIStorage UI.

The Longhorn dashboard opens.

4. Check the "in Progress" section of the dashboard to see whether there are any volumes stuck in the attaching/detaching state in. If a volume is that state, reboot its host.

- You may see the following error message during the Upgrade Cluster > Reapplying all settings > kubectl-patch :

```
kubectl rollout status deployment/rke2-ingress-nginx-controller -n kube-  
system --timeout=5m  
error: timed out waiting for the condition
```

If you see this error, do the following:

1. Check whether all the Kubernetes nodes are ready for scheduling. Run the following command from the ECS Server node:

```
kubectl get nodes
```

You will see output similar to the following:

```
NAME STATUS ROLES AGE VERSION
```

```
<node1> Ready,SchedulingDisabled control-plane,etcd,master 103m v1.21.11+rke2r1
<node2> Ready <none> 101m v1.21.11+rke2r1
<node3> Ready <none> 101m v1.21.11+rke2r1
<node4> Ready <none> 101m v1.21.11+rke2r1
```

2. Run the following command from the ECS Server node for the node showing a status of SchedulingDisabled:

```
kubectl uncordon
```

You will see output similar to the following:

```
<node1>node/<node1> uncordoned
```

3. Scale down and scale up the rke2-ingress-nginx-controller pod by running the following command on the ECS Server node:

```
kubectl delete pod rke2-ingress-nginx-controller-<pod number> -n kube-system
```

4. Resume the upgrade.

Related Information

[Backing up and restoring CDE jobs](#)