Cloudera on Premises Data Services 1.5.5

Cloudera Data Services on premises Release Notes

Date published: 2023-12-16
Date modified: 2025-06-06

https://docs.cloudera.com/

Legal Notice

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

Data Services Release Notes

Cloudera Data Services on premises 1.5.5 includes the Cloudera Management Console, Cloudera Data Catalog, Cloudera Data Warehouse, Cloudera AI, and Cloudera Data Engineering. Learn about the new features and improvements in each of these services.

Release notes for component services

- Data Catalog
- Management Console
- Cloudera Data Warehouse
- Cloudera AI
- Cloudera Data Engineering
- Cloudera Manager
- Replication Manager

What's new in Cloudera Data Services on premises 1.5.5

Understand the functionalities and improvements to features of Cloudera Control Plane install and upgrade components in Cloudera Data Services on premises 1.5.5.

Certificate Management

Cert-manager is an open-source tool for Kubernetes that automates the provisioning, management, and renewal of TLS certificates. Its documentation at https://cert-manager.io/docs/ provides comprehensive guidance on installing, configuring, and using cert-manager to secure workloads with trusted X.509 certificates. Cloudera provides out-of-the-box support for Venafi Trust Protection Platform (TPP) as part of the Cloudera Cloudera Embedded Container Service installation. By integrating cert-manager, the Data services achieve secure communication, reduced manual overhead, and compliance with security standards, leveraging its robust automation and flexibility. For more information on setting Cert-manager using Venafi TPP, see Setting up Certification Manager using Venafi TPP.

New upgrade prechecks

New pre-upgrade checks have been added to the list. The additional checks verify if the control plane and the docker registry is ready for upgrade. For more information, see Pre-upgrade checklist.

Quota Management for multiple base cluster support

Quota management enables you to control how resources are allocated within your Cloudera Data Services on premises clusters. In order to prevent a single workload from consuming all available cluster resources, you can limit the number of CPUs, GPUs, and memory allocated by application, user, business units, or Data Service by defining resource pools that define resource limits. Pools are organized in a hierarchical manner by defining nodes in the hierarchy with resource limits, which can then be subdivided as needed to allocate resources for an organization and to allocate resources to cluster or environment wide services such as the monitoring service. For information, see Quota Management.

Creating multiple environments with different base or Data Lake clusters

To register an environment with a data lake cluster managed by a Cloudera Manager that is different from your existing Cloudera Manager, you need to add the certificates of the new Cloudera Manager to the Cloudera Management Console UI. If the existing Cloudera Manager and the new Cloudera Manager share the same root CA, and the root CA is already uploaded as the data lake certificate, then no additional certificate needs to be added. For more information, see Creating multiple environments with different base or Data Lake clusters.

What's new in Platform Support 1.5.5

You must be aware of the platform support for Cloudera Data Services on premises 1.5.5.

Platform Certifications

- Cloudera Base on premises (7.1.9, 7.1.9 SP1 CHF5, 7.1.7 SP3 CHF10)
- Cloudera Manager 7.13.1 CHF 3
- Iceberg v2 GA on Cloudera Data Warehouse, Cloudera Data Engineering, & Cloudera AI with Ozone
- RHEL 8.10, 9.3, 9.4, 9.5
- OEL (RHCK Kernel Only) 8.10, 9.3, 9.4, 9.5
- Rocky Linux 8.10, 9.3, 9.4, 9.5
- K8s RKE2 1.30 and OCP 4.17

Note: Cloudera Manager 7.13.1 CHF 3 support Cloudera Data Services on premises 1.5.5 release.

Cloudera Manager 7.11.3 CHF8 does not support any Cloudera Data Services on premises release.

Repository Locations for 1.5.5

The URLs for Cloudera Data Services on premises 1.5.5 are listed in the following table:

URL Type	Repository Location
Index	`https://username:password@archive.cloudera.com/p/cdp-pvc-ds/1.5.5/`
Manifest	Repository: `https://username:password@archive.cloudera.com/p/cdp-pvc-ds/1.5.5/manifest.json`
Parcels	Repository: `https://username:password@archive.cloudera.com/p/cdp-pvc-ds/1.5.5/parcels/`

Fixed Issues for the Cloudera Data Services on premises 1.5.5

You can review the list of reported issues and their fixes in Cloudera Data Services on premises 1.5.5. Fixed issues represent selected issues that were previously logged through Cloudera Support, but are now addressed in the current Cloudera Data Services on premises release. These issues may have been reported in previous versions of Cloudera Data Services on premises as a known issue; meaning they were reported by customers or identified by Cloudera Quality Engineering teams.

OPSX-4308 - Display error in UI if listEnvironments failed

Error is now displayed on the Environments Page of the Cloudera Management Console UI, if an API failure is encountered.

OPSX-6048 - Clean up delete backup Custom Resource (CR) after the job is run

`DeleteBackup` now removes the backup deletion CR from resource deletebackuprequests.drs.cdp.cloudera.com

OPSX-5944 - Issues while uncordoning nodes during restart

The uncordon step was added into Cloudera Manager and is removed from the Cloudera Embedded Container Service parcel.

OPSX-5852 - Remove warn logs for "Unexpected partition in crn" from Cloudera Data Services on premises

"Unexpected partition in crn" log entries are now removed from the logs.

OPSX-5403 - Typecasting fails when truststore password is integer

When truststore password is set to all numbers (integer or float), control plane installation was failing in both Cloudera Embedded Container Service and OpenShift Container Platform. Safe datatype conversion is done to treat even numbers as string password. Even if numbers are used for truststore passwords, control plane installation will be successful.

OPSX-5903 - Upgrade failed with rke2-ingress-nginx-controller" exceeded its progress deadline

Automated the manual workaround of scaling down and scaling up the deployment when the earlier rollout or its status check fails.

OPSAPS-72270- ECS Restart]Start ECS| Start ECS command fails on uncordon nodes step

To resolve this issue:

1. Ensure the kube-apiserver is up and running for at least 60 seconds before proceeding with the uncordon step.
2. Use the correct target node name, not just the name of the node where the uncordon command is executed.

Known issues for the Cloudera Data Services on premises 1.5.5

You must be aware of the known issues and limitations, the areas of impact, and workaround in Cloudera Data Services on premises 1.5.5 release.

Known Issues in Cloudera Data Services on premises 1.5.5

OBS-8038: When using the Grafana Dashboard URL shortener, the shortened URL defaults to localhost:3000. This behaviour happens because the URL shortener uses the local server address instead of the actual domain name of the Cloudera Observability instance. As a result, users cannot access the shortened URL.

You must not use the shortened URL. To ensure users can access the URL, update it to use the correct Cloudera Observability instance domain name, such as cp_domain/{shorten_url}{}.

Note: cp_domain refers to the Cloudera Control Plane domain.

OPSX-6209 and DWX-20809: Cloudera Data Services on premises installations on RHEL 8.9 or lower versions may encounter issues

You may notice issues when installing Cloudera Data Services on premises on Cloudera Embedded Container Service clusters running on RHEL 8.9 or lower versions. Pod crashloops are noticed with the following error:

```
Warning  FailedCreatePodSandBox           1s (x2 over 4s)  kubel
et   Failed to create pod
```

```
sandbox: rpc error: code = Unknown desc = failed to create contai
nerd task: failed to create
shim task: OCI runtime create
 failed: runc create failed: unable to start container process: u
nable to init seccomp: error
loading seccomp filter into kernel: error loading seccomp filt
er: errno 524: unknown
```

The issue is due to a memory leak with 'seccomp' (Secure Computing Mode) in the Linux kernel. If your kernel version is not on 6.2 or higher verisons or if it is not part of the list of versions mentioned here, you may face issues during installation.

To avoid this issue, increase the value of net.core.bpf_jit_limit by running the following command on all ECS hosts:

```
[root@host ~]# sysctl net.core.bpf_jit_limit=528482304
```

However, Cloudera recommends upgrading the Linux kernel to an appropriate version that contains a patch for the memory leak issue. For a list of versions that contain this patch, see this link.

Note: RHEL 9.0, 9.1, and 9.2 will not be supported in 1.5.5 because of https://bugzilla.redhat.com/show_bug.cgi?id=2218682. You will need to upgrade to at least RHEL 9.3 if you are on 9.x line.

COMPX-20705: [153CHF-155] Post ECS upgrade pods are stuck in ApplicationRejected State

After upgrading the CDP installation pods on Kubernetes could be left in a failure state showing "ApplicationRejected". This is caused by a delay in settings being applied to Kubernetes as part of the post upgrade steps.

To resove this issue, restart the scheduler to pick up the latest settings for Kubernetes. Also, restart YuniKorn using the following commands:

```

kubectl scale deployment yunikorn-scheduler --replicas=0 -n yun
ikorn
kubectl scale deployment yunikorn-scheduler --replicas=1 -n yu
nikorn
```

OPSX-6303 - ECS server went down - 'etcdserver: mvcc: database space exceeded'

ECS server may fail with error message - "etcdserver: mvcc: database space exceeded" in large clusters.

1. Add to the safety valve for server group:

   ```

   etcd-arg:
   - "quota-backend-bytes=4294967296"
   ```
2. Restart stale services (Select the option `re-deploy client configs`).
3. The default value for `quota-backend-bytes` is 2 GB. It can be increased up to 8 GB.

OPSX-6295 - Control Plane upgrade failing with cadence-matching and cadence-history

Incase of extra cadence-matching and cadence-history pod stuck in Init:CreateContainerError state , Cloudera Embedded Container Service Upgrade to 1.5.5 will be stuck in retry loop because of all pods running validation failure.

You need to manually apply the workaround to proceed further upgrade and get it done successfully. Hence, delete the stuck cadence pods.

OPSX-4391 - External docker cert not base64 encoded

When using Cloudera Data Services on premises on ECS, in some rare situations, the CA certificate for the Docker registry in the cdp namespace is incorrectly encoded, resulting in TLS errors when connecting to the Docker registry.

Compare and edit the contents of the "cdp-private-installer-docker-cert" secret in the cdp namespace so that it matches the contents of the "cdp-private-installer-docker-cert" secret in other namespaces. The secrets and their corresponding namespaces can be identified using the command:

```
kubectl get secret -A | grep cdp-private-installer-docker-cert
```

Inspect each secret using the command:

```
kubectl get secret -n cdp cdp-private-installer-docker-cert -o y
aml
```

Replace "cdp" with the different namespace names. If necessary, modify the secret in the cdp namespace using the command:

```
kubectl edit secret -n cdp cdp-private-installer-docker-cert
```

OPSX-6245 - Airgap | Multiple pods are in pending state on rolling restart

Performing back-to-back rolling restarts on ECS clusters can intermittently fail during the Vault unseal step. During rapid consecutive rolling restarts, the kube-controller-manager pod may not return to a ready state promptly. This can cause a cascading effect where other critical pods, including Vault, fails to initialize properly. As a result, the unseal Vault step fails.

As a workaround, perform the following steps:

1. Stop the ECS role that failed.
2. Start the ECS role again.
3. If required, perform the rolling restart again.

OPSX-4684 - Start ECS command shows green(finished) even though start docker server failed on one of the hosts

Docker service starts with one or more docker roles failed to start because the corresponding host is unhealthy.

Make sure the host is healthy. Start the the docker role in the host.

OPSX-5986 - ECS fresh install failing with helm-install-rke2-ingress-nginx pod failing to come into Completed state

ECS fresh install fails at the "Execute command Reapply All Settings to Cluster on service ECS" step due to a timeout waiting for helm-install.

To confirm the issue, run the following kubectl command on the ECS server host to check if the pod is stuck in a running state:

```
kubectl get pods -n kube-system | grep helm-install-rke2-ingress-
nginx
```

To resolve the issue, manually delete the pod by running:

```
kubectl delete pod <helm-install-rke2-ingress-nginx-pod-name> -n
 kube-system
```

Then, click Resume to proceed with the fresh install process on the Cloudera Manager UI.

OPSX-6298 - Issue on service namespace cleanup

There might be cases in which uninstalling services from the Cloudera Data Services on premises UI will fail due to various reasons.

In case uninstallation of a Service fails, trigger again the service uninstall process, and mark “Force Delete” to ensure that all metadata of the service will be removed from Cloudera side. Then, move to the OpenShift UI, and there search for that service namespace / project. On that project/namespace select the Action button on the top right of the screen and choose to Delete the Project.

If you move back to the main Project screen you could see that the project is moving to status “Terminating” after which it will be removed from the OCP platform. That action will ensure that all the entities linked to that project/namespace will also be removed by OpenShift.

OPSX-6265 - Setting inotify max_user_instances config

We cannot recommend an exact value for inotify max_user_instances config. It depends on all workloads that are run in a given node.

With newly introduced features like istio, cert manager, in Cloudera Control Plane, there is a need to set inotify max_user_instancesconfig to 256 instead of 128 to resolve this issue.

COMPX-20362 - Use API to create a pool that has a subset of resource types

The Resource Management UI supports displaying only three resource types: CPU, memory and GPU. The Resource Management UI will always set all three resource types it knows about: CPU, Memory and GPU (K8s resource nvidia.com/gpu) when creating a quota. If no value is chosen for a resource type a value of 0 will be set, blocking the use of that resource type.

To create a pool that has a subset of resource types the REST API must be used as follows:

```
POST /api/v1/compute/createResourcePool
```

```
Payload:



{
    "pool": {
        "path": "root.environment.service.mypool",
        "policy": {
            "allocation": "INELASTIC"
        },
        "quota": {
            "cpu": "100 m",
            "memory": "10 GB"
        }
    }
}
```

Note: Payload needs to be confirmed and checked.

Known issues from previous releases carried in Cloudera Data Services on premises 1.5.5

Known Issues identified in 1.5.4

DOCS-21833: Orphaned replicas/pods are not getting auto cleaned up leading to volume fill-up issues

By default, Longhorn will not automatically delete the orphaned replica directory. One can enable the automatic deletion by setting orphan-auto-deletion to true.

No workaround available.

OPSX-5310: Longhorn engine images were not deployed on ECS server nodes

Longhorn engine images were not deployed on ECS server nodes due to missing tolerations for Cloudera Control Plane taints. This caused the engine DaemonSet to schedule only on ECS agent nodes, preventing deployment on Cloudera Control Plane nodes.

1. Check the Engine `DaemonSet Status`. Run the following command to check if the Longhorn engine DaemonSet is missing on certain nodes:

   ```
   kubectl get ds -n longhorn-system | grep engine
   ```

2. Identify Taints on Affected Nodes. Run the following command to check for taints on affected nodes:

   ```
   kubectl describe node <node-name> | grep Taints
   ```

   Note: If you see, `node-role.kubernetes.io/control-plane=true:NoSchedule`, this confirms the issue.

3. Manually Edit the DaemonSet to Add a Toleration. Edit the Longhorn engine DaemonSet YAML:

   ```
   kubectl edit ds -n longhorn-system engine-image-ei-<your-eng
   ine-id>
   ```

4. Add the following under tolerations:

   ```

   tolerations:
   - effect: NoSchedule
     key: node-role.kubernetes.io/control-plane
     operator: Equal
     value: "true"
   ```

5. Apply the changes and verify deeployment. Save and exit the editor. Then, check if the DaemonSet is now running on all necessary nodes:

   ```
   kubectl get pods -n longhorn-system -o wide | grep engine
   ```

   Verify that the engine pods are successfully scheduled on the affected ECS server nodes.

OPSX-5155: OS Upgrade | Pods are not starting after the OS upgrade from RHEL 8.6 to 8.8

After an OS upgrade and start of the Cloudera Embedded Container Service service, pods fail to come up due to stale state.

Restart the Cloudera Embedded Container Service cluster.

OPSX-5055: Cloudera Embedded Container Service upgrade failed at Unseal Vault step

During an Cloudera Embedded Container Service upgrade from 1.5.2 to 1.5.4 release, the vault pod fails to start due to an error caused by the Longhorn volume unable to attach to the host. The error is as below:

Warning FailedAttachVolume 3m16s (x166 over 5h26m) attachdetach-controller AttachVolume.Attach failed for volume "pvc-0ba86385-9064-4ef9-9019-71976b4902a5" : rpc error: code = Internal desc = volume pvc-0ba86385-9064-4ef9-9019-71976b4902a5 failed to attach to node host-1.cloudera.com with attachmentID csi-7659ab0e6655d308d2316536269de47b4e66062539f135bf6012bfc8b41fc345: the volume is currently attached to different node host-2.cloudera.com

Follow below steps provided by SUSE to ensure the Longhorn volume is correctly attached to the node where the vault pod is running.

```

```

```
# Find out the volume name that is failing to attach to the vault
 pod.
For e.g. pvc-bc73e7d3-c7e7-468a-b8e0-afdb8033e40b from the pod
logs.
kubectl edit volumeattachments.longhorn.io -n longhorn-system
pvc-bc73e7d3-c7e7-468a-b8e0-afdb8033e40b

# Update the "spec:" section of the volumeattachment and replace
attachmentTickets section with {} as shown below and save.
spec:
 attachmentTickets: {}
 volume: pvc-bc73e7d3-c7e7-468a-b8e0-afdb8033e40b

# scale down the vault statefulset to 0 and scale it back up.
kubectl scale sts vault --replicas=0 -n vault-system
kubectl scale sts vault --replicas=1 -n vault-system
```

OPSX-4684: Start Cloudera Embedded Container Service command shows green(finished) even though start docker server failed on one of the hosts

The Docker service starts, but one or more Docker roles fail to start because the corresponding host is unhealthy.

Ensure the host is healthy. Start the the Docker role on the host.

OPSX-735: Kerberos service should handle Cloudera Manager downtime

The Cloudera Manager Server in the base cluster operates to generate Kerberos principals for Cloudera on premises. If there is downtime, you may observe Kerberos-related errors.

Resolve downtime on Cloudera Manager. If you encounter Kerberos errors, you can retry the operation (such as retrying creation of the Virtual Warehouse).

Known Issues identified in 1.5.2

OPSX-4594: [ECS Restart Stability] Post rolling restart few volumes are in detached state (vault being one of them)

After rolling restart there may be some volumes in detached state.

1. Open the Longhorn UI to view the detached volumes.
2. Perform the following operations for each volume in a detached state:

   a. Identify the workload name and type from the volume details.
   b. Identify the workload and number of replicas using kubectl or the Kubernetes UI.
   c. Scale the workload down to 0.
   d. Wait for the pods associated with the workload to fully terminate.
   e. Scale up the workload up to the number of replicas it had originally.

To prevent this issue, use the Longhorn UI to set the number of replicas for the volume to at least 3.

OPSX-4392: Getting the real client IP address in the application

CML has a feature for adding the audit event for each user action (Monitoring User Events). In Private Cloud, instead of the client IP, we are getting the internal IP, which is logged into the internal DB.

In ECS, add the enable-real-ip configuration as true for the nginx ingress controller:

```
apiVersion: v1
data:
  allow-snippet-annotations: "true"
  enable-real-ip: "true"                   <<<<<<<<<< new config
kind: ConfigMap
metadata:
  annotations:
    meta.helm.sh/release-name: rke2-ingress-nginx
    meta.helm.sh/release-namespace: kube-system
  creationTimestamp: "2023-05-09T04:54:53Z"
  labels:
    app.kubernetes.io/component: controller
    app.kubernetes.io/instance: rke2-ingress-nginx
    app.kubernetes.io/managed-by: Helm
    app.kubernetes.io/name: rke2-ingress-nginx
    app.kubernetes.io/part-of: rke2-ingress-nginx
    app.kubernetes.io/version: 1.6.4
    helm.sh/chart: rke2-ingress-nginx-4.5.201
  name: rke2-ingress-nginx-controller
  namespace: kube-system
  resourceVersion: "162559439"
  uid: cca67b0c-bc05-4e1f-8439-7d44323f4624
```

In OpenShift Container Platform, you may be able configure this using HAproxy with X-forward-for pass to OpenShift 4.

CDPVC-1137, CDPAM-4388, COMPX-15083, and COMPX-15418: OpenShift Container Platform version upgrade from 4.10 to 4.11 fails due to a Pod Disruption Budget (PDB) issue

PDB can prevent a node from draining which makes the nodes to report the "Ready,SchedulingDisabled" state. As a result, the node is not updated to correct the Kubernetes version when you upgrade OpenShift Container Platform from 4.10 to 4.11.

To resolve this issue, confirm that the upgrade has failed due to the PDB issue, and then manually delete the PDBs from the Cloudera on premises namespace.

1. Run the following command to check whether the nodes are stuck in the "Ready,SchedulingDisabled" state:

   ```
   oc get nodes
   ```

2. Get the machine config daemon details of the particular pod as follows:

   ```
   oc get po -n openshift-machine-config-operator -l 'k8s-app=m
   achine-config-daemon' -o wide
   ```

3. Check the logs of the machine config operator of that particular node as follows:

```
oc logs -f -n openshift-machine-config-operator [***MACHINE-CONFIG-DAEMON-NAME***] -c machine-config-daemon
```

Replace *[***MACHINE-CONFIG-DAEMON-NAME***]* with the actual machine config daemon name.

You may see one of the following errors in the node logs:

- error when evicting pods/cdp-release-cpx-liftie-****" -n "*[***PRIVATE-CLOUD-NAMESPACE***]* Cannot evict pod as it would violate the pod's disruption budget
- error when evicting pods/"cdp-release-cluster-proxy-[******]" -n "*[***PRIVATE-CLOUD-NAMESPACE***]* Cannot evict pod as it would violate the pod's disruption budget

Delete the PDB from the Cloudera on premises namespace as follows:

a. Obtain the PDB for the cdp-release-cluster-proxy namespace:

```
oc get pdb -n [***PRIVATE-CLOUD-NAMESPACE***] | grep cdp-release-cluster-proxy
```

b. Back up the PDB:

```
oc get pdb [***PDB-NAME-OF-CLUSTER-PROXY***] -n [***PRIVATE-CLOUD-NAMESPACE***] -o yaml >> [***BACKUP-FILE-NAME***].yaml
```

c. Delete the PDB:

```
oc delete pdb [***PDB-NAME-OF-CLUSTER-PROXY***] -n [***PRIVATE-CLOUD-NAMESPACE***]
```

Repeat the steps to delete the cdp-release-cpx-liftie PDB as well.