

Repurposing CDP Private Cloud Base Nodes for CDP Private Cloud Data Services on ECS

Date published: 2023-12-16

Date modified: 2024-03-23

CLouDERA

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

Repurposing CDP Private Cloud Base Nodes for CDP Private Cloud Data Services on ECS.....	4
---	----------

Repurposing CDP Private Cloud Base Nodes for CDP Private Cloud Data Services on ECS

Separating compute and storage using CDP Private Cloud (PvC) Data Services provides cloud-native capabilities to help optimize compute resources.

About this task

Workloads that are candidates for migration from CDP PvC Base to CDP PvC Data Services include CDSW, Hive, Spark on YARN, and Impala. You can identify underutilized nodes in CDP PvC Base, repurpose them as PvC Data Services nodes, and then migrate existing workloads from PvC Base to PvC Data Services.

Before you begin

Review the following hardware requirements before repurposing the PvC Base nodes for PvC Data Services.

1. Check the [CDP Private Cloud Data Services Hardware Requirements](#) and confirm that the nodes you intend to repurpose meet the requirements. This may include new requirements that necessitate added storage such as fast cache devices, increased RAM, or upgraded network cards.
2. Ensure that your existing PvC Base cluster complies with [PvC Data Services system requirements](#). If not, you may need to [upgrade your Private Cloud Base cluster](#) to a supported version.
3. Ensure that your existing PvC Base cluster will still function properly without the repurposed nodes. Remember that removing nodes reduces overall storage and compute capacity in the PvC Base cluster.
 - Target existing worker nodes – avoid repurposing “master” nodes (i.e., the NameNode).
 - Avoid using edge nodes, or gateways.
 - Calculate how much storage is being removed (the combined footprint of the JBOD drives on nodes). The revised utilized capacity for HDFS, Ozone, and Kudu should generally not exceed 75% after the nodes are removed.
 - Review the [PvC Base cluster utilization reports](#) and ensure that the remaining compute capacity can support the remaining PvC Base workloads:
 - a. HBase, Solr, MapReduce, and NiFi workloads will remain on PvC Base. Ensure that your remaining compute capacity can support these workloads.
 - b. Ranger, Atlas, Ozone, HMS, Zookeeper, and HDFS will also remain on PvC Base. Ensure that your remaining compute capacity can support these workloads.

Repurposing the nodes

If the high availability (HA) configuration is not changing on the PvC Base cluster, you can run the following steps to repurpose nodes without any downtime.

1. [Remove the host](#) from the PvC Base cluster. Using the Cloudera Manager Hosts Decommission feature ensures that workloads are shut down gracefully on each node, and no new work is accepted.
 - When the target includes a HDFS Datanode role, HDFS blocks will be evicted from that host and copied to available space in the cluster. This can take some time, depending on how much data is stored and how your environment is set up.
 - After removing a host from the PvC Base cluster, it is highly recommended that you treat this node according to your normal procedures for redeployed hardware. This may involve wiping drives and reimaging the operating system (OS).
2. Treat the old node as a new server, commissioning it by adding it to an existing PvC Data Services Cluster or including it in a new PvC Data Services cluster. See [Install CDP Private Cloud Data Services](#) for more information.

Related Information

[Migrating Data Science Workbench to Machine Learning](#)

[Migrating Hive workloads from CDP Private Cloud Base to CDW Private Cloud](#)

[Moving Impala compute workloads from CDP Private Cloud Base to CDW Data Service on Private Cloud](#)

[Repurposing CDSW nodes for CML](#)