

# Disk Vertical Scaling in AWS and Azure (Preview)

Date published: 2023-12-12

Date modified: 2024-04-18

## Legal Notice

© Cloudera Inc. 2024, 2023. 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.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Contents

<b>About this feature</b>	<b>4</b>
Introduction	4
Limitations	4
AWS permissions	5
Azure permissions	5
<b>Download Beta CDP CLI</b>	<b>5</b>
<b>Modify Disks</b>	<b>6</b>
Modifying disks using CDP CLI	6
Verifying that modification is complete	6
Verifying disk size/type in CDP UI	7
Verifying disk size/type in AWS Console	8
<b>Add Disks</b>	<b>8</b>
Adding disks using CDP CLI	8
<b>Deleting Disks</b>	<b>9</b>
Deleting disks using CDP CLI	9

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

# About this feature

## Introduction

There are many clusters that are deployed with standard magnetic storage. With the growing lineage data, these disks are running out of space on core nodes. These need to be moved to General Purpose SSDs (gp2/gp3 on AWS) and or resized to a bigger disk.

The disks attached to the Data Lake and Data Hub clusters can be changed or resized in AWS and Azure without downtime.

## Limitations

When using this preview feature, be aware of the following limitations:

- This feature is only available for AWS and AZURE.
- The disks can only be resized up, meaning you cannot reduce the size of an attached block storage. If there are multiple disks of different sizes, the size of all the disks attached to the instances in the group that are smaller or lesser than the requested size will be increased to the requested size.
- This feature will only resize additional block storages in an instance and not the root volume.
- Clusters and cluster services must be in running state before disk vertical scaling is performed.
- Current implementation does not support this feature through CDP UI; It is available only through Beta CDP CLI.

In addition, the following limitations apply to this feature when it is used on AWS:

- The disk modification feature on AWS can only be used once in 6 hours. This is a limitation on the AWS side.

In addition, the following limitations apply to this feature when it is used on Azure:

- Disk type change is not supported for AZURE.
- Not supported for Ultra disks or Premium SSD v2 disks.
- Not supported for shared disks.
- Not available on some classic VMs.
- If a disk is 4 TB or less, you should deallocate your VM and detach the disk before expanding it beyond 4 TB. If a disk is already greater than 4 TB, you can expand it without deallocating the VM and detaching the disk, as mentioned in [Expand virtual hard disks on a Linux VM](#).
- If the above points aren't satisfied, resizing will require downtime; That is, CDP will have to deallocate (stop VM), and then resize and start VM.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## AWS permissions

This feature requires the following permissions to be added to the cross-account policy described in [Cross-account access IAM role](#):

- ec2:ModifyVolume
- ec2:DescribeVolumesModifications
- ec2:DescribeVolumeStatus

The following table explains why CDP needs these permissions:

Permission	Description
ec2:ModifyVolume	ModifyVolume is required to modify the volume attributes such as type, size and IOPS capacity. Without this, volume modifications cannot be performed by CDP.
ec2:DescribeVolumesModifications	This is required to verify whether the volume modifications performed by CDP were successful. Only upon successful modification, other steps like resizing will be done.
ec2:DescribeVolumeStatus	This is required to make sure that the volume being modified is attached to an instance and not an orphaned volume.

## Azure permissions

This feature requires the following permissions to be added to the minimal set of permissions described in [Prerequisites for the provisioning credential](#):

- Microsoft.Compute/disks/write

This permission allows users to update disk sizes on a resource.

## Download Beta CDP CLI

You have to use Beta CDP CLI to modify disks attached to an instance group. To install Beta CDP, refer to [Installing Beta CDP CLI](#).

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

# Modify Disks

## Modifying disks using CDP CLI

Use the following Beta CDP CLI command to modify the volumes attached to the instances in a host group. Replace the placeholders with actual values. For example <DATA LAKE-NAME> should be replaced with an actual name. As part of this update, the instance-template parameter in the vertical scaling command has been made optional. But one of the instance-template or disk-options have to be provided.

Additional parameters for vertical scaling have been added to both the datalake and datahub commands. The modification request is sent as part of the `--disk-options` parameter.

The <VOLUME-TYPE> placeholder is for the type of volume the disks are being modified to. It is an optional field and should be added if the volume type has to be modified, in which case the type will not be modified.

The <SIZE> placeholder is optional as well and is for the size the disks are being increased to in GB. Please specify only the integer value.

```
//DATA LAKE
cdp datalake start-datalake-vertical-scaling
    --datalake <DATA LAKE-NAME>
    --group <INSTANCE-GROUP-NAME>
    --disk-options
modifyDisks="{volumeType=\"<VOLUME-TYPE>\", size=<SIZE>}"

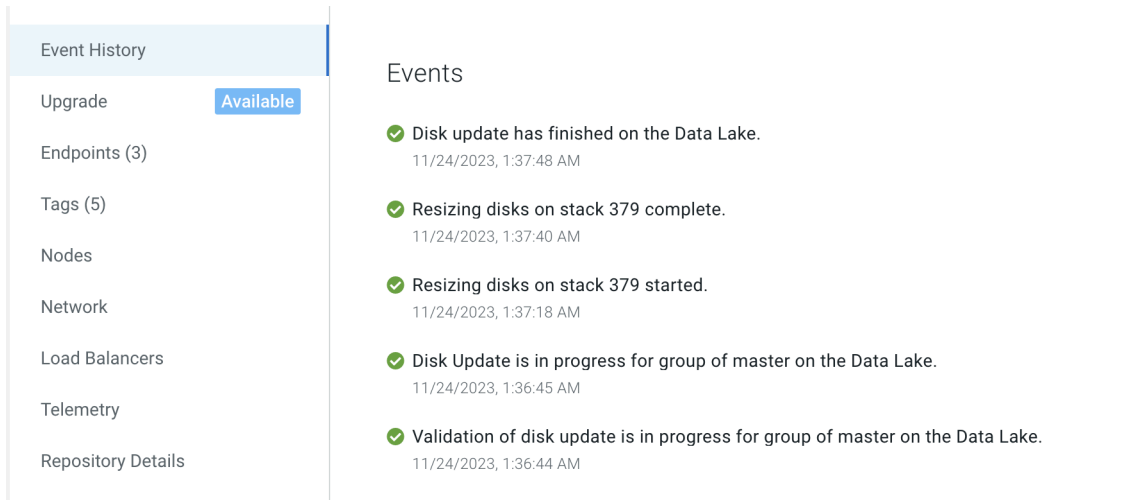
//DATA HUB
cdp datahub start-cluster-vertical-scaling
    --datahub <DATAHUB-NAME>
    --group <INSTANCE-GROUP-NAME>
    --disk-options
modifyDisks="{volumeType=\"<VOLUME-TYPE>\", size=<SIZE>}"
```

## Verifying that modification is complete

The change to the disk can be verified through CDP UI, AWS console, or by logging into the instances directly, after the flow is completed in the Event History.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

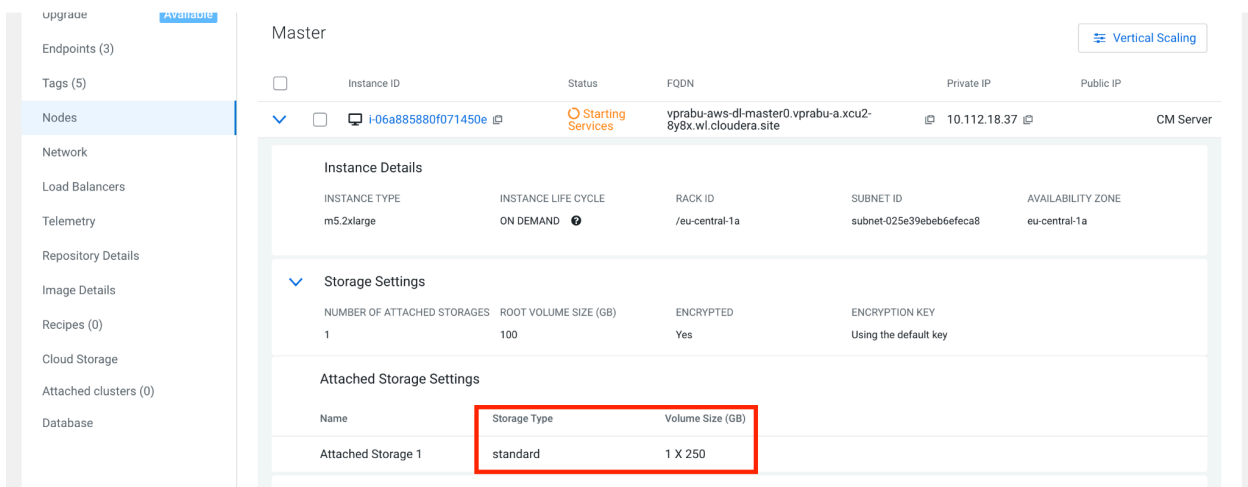
Here is a screenshot from Event History showing the completion of the disk update:



## Verifying disk size/type in CDP UI

1. Navigate to the **Management Console > Data Lake** or **Management Console > Data Hub**.
2. Click into the Data Lake / Data Hub that was modified.
3. Click on the **Nodes** in the left hand tree.
4. Open accordion of the instance group that was modified.
5. Open **Storage Settings** accordion in any of the instances in the group.
6. The Storage Type and Volume Size should be updated based on the request.

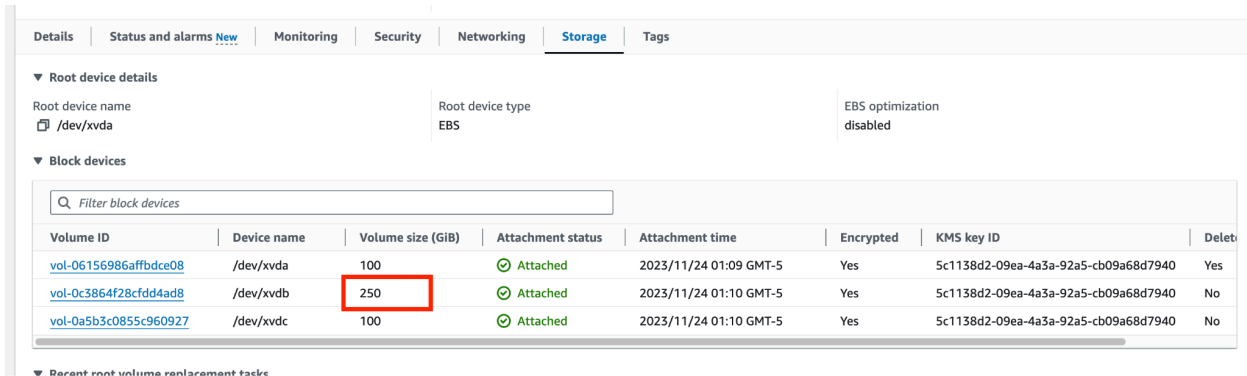
Environments / vprabu-aws-resize1 / Data Lake / Nodes



*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Verifying disk size/type in AWS Console

1. Navigate to the **Management Console > Data Lake** or **Management Console > Data Hub**.
2. Click into the Data Lake / Data Hub that was modified.
3. Click on the **Nodes** in the left hand tree.
4. Click on the AWS link for the instance and log into AWS Console.
5. Select the Storage tab for the instance.
6. The volume size should be updated as per the request.
7. You can also click into an individual volume to make sure the type and size are modified correctly.



## Add Disks

In some cases when customers run out of memory, it is cheaper to add an additional volume instead of resizing to a bigger volume. This feature allows for adding additional block storages to an instance group with minimum interruption. Only the services running on the instance group are stopped, as the additional volume has to be mounted and configured for use on the instances.

Based on the request, ‘N’ number of disks of the same type and size can be added to the instance group.

## Adding disks using CDP CLI

Use the following CDP CLI Beta command to add additional volumes attached to the instances in a host group. Replace the placeholders with actual values. For example <DATA LAKE-NAME> should be replaced with an actual name. As part of this update, the instance-template parameter in the vertical scaling command has been made optional. But one of the instance-template or disk-options have to be provided.

The add disks request is sent as part of the `--disk-options` parameter. All parameters for the add disk input are required.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*



The <NUMBER\_OF\_DISKS> placeholder is for the number of volumes being added.

The <VOLUME-TYPE> placeholder is for the type of volume being added.

The <SIZE> placeholder is for the size the disks being added in GB. Please specify only the integer value.

The cloudVolumeUsageType is the purpose for which the disk is being added - it is an enum field with either "GENERAL" or "DATABASE" value.

```
//DATALAKE

cdp datalake start-datalake-vertical-scaling --datalake=<DATALAKE_NAME>
--group=<DATALAKE_INSTANCE_GROUP_NAME> --disk-options
addDisks="{numberOfDisks=<NUMBER_OF_DISKS_TO_ADD>,
volumeType=\"<TYPE_OF_VOLUME_TO_ADD>\",size=<SIZE_OF_VOLUME>,cloudVolume
UsageType=\"GENERAL\"|\"DATABASE\"}"

//DATAHUB

cdp datahub start-cluster-vertical-scaling --datahub=<DATAHUB_NAME>
--group=<DATAHUB_INSTANCE_GROUP_NAME> --disk-options
addDisks="{numberOfDisks=<NUMBER_OF_DISKS_TO_ADD>,
volumeType=\"<TYPE_OF_VOLUME_TO_ADD>\",size=<SIZE_OF_VOLUME>,cloudV
olumeUsageType=\"GENERAL\"|\"DATABASE\"}"
```

## Deleting Disks

In cases where only compute services are being run on an instance group, any block storage attached to the instance is going to cost the customer even if the cluster is stopped. As compute services do not store any persistent data on attached volumes, having additional volumes to store temporary data is unnecessary. In such cases, customers can use this command to delete all attached volumes on instances in an instance group.

This can be done only for compute instance groups on only datahubs, as datalakes need persistent volumes. This command deletes all additional volumes for instances in a instance group.

## Deleting disks using CDP CLI

Use the following CDP CLI Beta command to deleting all attached additional volumes to the instances in a host group. Replace the placeholders with actual values. For example <DATALAKE-NAME> should be replaced with an actual name. As part of this update, the instance-template parameter in the vertical scaling command has been made optional. But one of the instance-template or disk-options have to be provided.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## CLOUDERA TECHNICAL PREVIEW DOCUMENTATION

The delete disks request is sent as part of the `--disk-options` parameter. All parameters for the add disk input are required.

The `<deleteDisks>` parameter accepts a boolean value with `true/ false`.

```
//DATAHUB  
  
cdp datahub start-cluster-vertical-scaling --datahub=<DATAHUB_NAME>  
--group=<DATAHUB_INSTANCE_GROUP_NAME> --disk-options  
deleteDisks=true|false
```