

Disk Vertical Scaling - Disk Type Change and Resizing in AWS (Preview)

Date published: 12/12/23

Date modified: 12/12/23

Contents

Contents	2
About this feature	3
Introduction	3
Limitations	3
AWS permissions	3
Download Beta CDP CLI	4
Modifying Disks	4
Modify disks using the beta CDP CLI	4
Verifying modification is complete	5
Verifying disk size/type in CDP UI	5
Verifying disk size/type in AWS Console	6

About this feature

Introduction

Many AWS clusters are deployed with standard magnetic storage. With growing lineage data, these disks can run out of space on core nodes. If necessary, you can move these disks to General Purpose (gp2/gp3) SSDs and/or resize to a bigger disk.

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

Limitations

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

- This feature is only available for AWS. Support for AZURE will be rolled out soon.
- 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.
- The disk modification feature on AWS can only be used once in 6 hours. This is a limitation on the AWS side.
- Current implementation does not support this feature through CDP UI. You must use the Beta CDP CLI.

AWS permissions

This feature requires the following permissions to be added to the Cross-Account policy:

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

The following table explains why CDP needs these permissions:

Permission	Description
<code>ec2:ModifyVolume</code>	<code>ModifyVolume</code> 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 can other steps like resizing take place.
ec2:DescribeVolumeStatus	This is required to make sure that the volume being modified is attached to an instance and not an orphaned volume.

Download Beta CDP CLI

You must use the beta CDP CLI to modify disks attached to an instance group.

1. Download the latest Beta CDP CLI. For more information see [Installing Beta CDP CLI](#).

```
pip install cpdcli-beta
```

If you have previously installed Beta CDP CLI, you can update it to the latest version using:

```
pip3 install cpdcli-beta --upgrade --user
```

Modifying disks

Modify disks using the beta CDP CLI

Use the following CDP CLI Beta 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 either `instance-template` or `disk-options` must be provided.

An additional parameter for vertical scaling has been added to both Data Lakes and Data Hubs. 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 a non-null field and can contain current disk type value, in which case the type will not be modified.

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

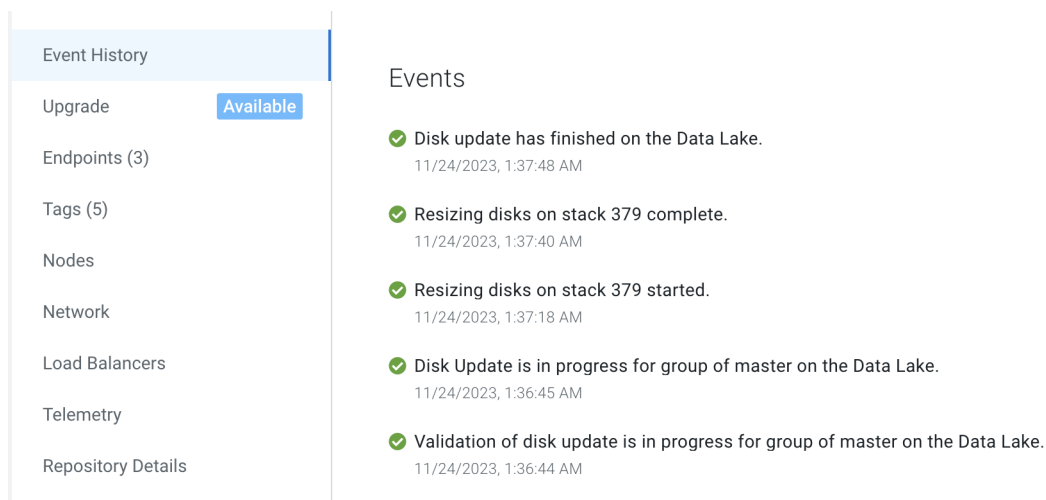
The parameters important for this feature are highlighted:

```
//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 <DATA HUB-NAME>
    --group <INSTANCE-GROUP-NAME>
    --disk-options
modifyDisks={volumeType=<VOLUME-TYPE>,size=<SIZE>}
```

Verifying modification is complete

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



Event	Timestamp
✓ Disk update has finished on the Data Lake.	11/24/2023, 1:37:48 AM
✓ Resizing disks on stack 379 complete.	11/24/2023, 1:37:40 AM
✓ Resizing disks on stack 379 started.	11/24/2023, 1:37:18 AM
✓ Disk Update is in progress for group of master on the Data Lake.	11/24/2023, 1:36:45 AM
✓ Validation of disk update is in progress for group of master on the Data Lake.	11/24/2023, 1:36:44 AM

Verifying disk size/type in CDP UI

1. Navigate to the **Management Console > Data Lake/Data Hub**.
2. Click into the Data Lake/Data Hub that was modified.
3. Click on **Nodes** in the left-hand menu.
4. Click the accordion button of the instance group that was modified.

5. Click the **Storage Settings** accordion button 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

upgrade **AVAILABLE** Vertical Scaling

Endpoints (3)
Tags (5)
Nodes
Network
Load Balancers
Telemetry
Repository Details
Image Details
Recipes (0)
Cloud Storage
Attached clusters (0)
Database

Master

Instance ID: I-06a885880f071450e Status: Starting Services FQDN: vprabu-aws-dl-master0.vprabu-a.xcu2-8y8x.wl.cloudera.site Private IP: 10.112.18.37 Public IP: CM Server

Instance Details

INSTANCE TYPE	INSTANCE LIFE CYCLE	RACK ID	SUBNET ID	AVAILABILITY ZONE
m5.2xlarge	ON DEMAND	/eu-central-1a	subnet-025e39ebeb6efeca8	eu-central-1a

Storage Settings

NUMBER OF ATTACHED STORAGES	ROOT VOLUME SIZE (GB)	ENCRYPTED	ENCRYPTION KEY
1	100	Yes	Using the default key

Attached Storage Settings

Name	Storage Type	Volume Size (GB)
Attached Storage 1	standard	1 X 250

Verifying disk size/type in AWS Console

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

Details | Status and alarms **New** | Monitoring | Security | Networking | **Storage** | Tags

▼ Root device details

Root device name: /dev/xvda Root device type: EBS EBS optimization: disabled

▼ Block devices

Filter block devices

Volume ID	Device name	Volume size (GiB)	Attachment status	Attachment time	Encrypted	KMS key ID	Delete
vol-06156986affbdce08	/dev/xvda	100	Attached	2023/11/24 01:09 GMT-5	Yes	5c1138d2-09ea-4a3a-92a5-cb09a68d7940	Yes
vol-0c3864f28cfd44d8	/dev/xvdb	250	Attached	2023/11/24 01:10 GMT-5	Yes	5c1138d2-09ea-4a3a-92a5-cb09a68d7940	No
vol-0a5b3c0855c960927	/dev/xvdc	100	Attached	2023/11/24 01:10 GMT-5	Yes	5c1138d2-09ea-4a3a-92a5-cb09a68d7940	No

▼ Recent root volume replacement tasks