

# CML Workspace Backup & Restore (Preview)

Date published: 2022-02-10

Date modified: 2022-05-09

## Legal Notice

© Cloudera Inc. 2022. 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.

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>Legal Notice</b>	<b>2</b>
<b>Description</b>	<b>4</b>
<b>Limitations</b>	<b>4</b>
Cloud Platform	4
CDP Environment	5
CML Workspace Endpoint	5
CLI Only User Interface	5
CML Version Compatibility	5
<b>Prerequisites</b>	<b>5</b>
CDP Entitlement	5
AWS Backup Service Opt-in	5
AWS CDP Cross-Account Role Permissions	6
Backup and Restore Permission IAM Policies	6
Backup Policy	6
Restore Policy	8
Trust Relationship	10
Attach Backup and Restore Policies to the Cross-Account Role	10
<b>How to Backup a CML Workspace</b>	<b>10</b>
<b>Backup Catalog Commands</b>	<b>11</b>
List Backup Instances of a Workspace	11
Delete Backup Instances of a Workspace	12
<b>How to Restore a CML Workspace</b>	<b>12</b>
<b>Known Issues</b>	<b>14</b>

*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.*

# CML Workspace Backup & Restore

## Tech Preview Documentation

### Description

Cloudera Machine Learning makes it easy to create machine learning projects, jobs, experiments, ML models, and applications in workspaces. The data and metadata of these artifacts are stored in different types of storage systems in the cloud. The Backup and Restore feature gives you the ability to backup all of your data to protect your machine learning artifacts against disasters. If your Cloudera Machine Learning workspace is backed up, this feature lets you restore the saved data into a new CML workspace so that you can recover your ML artifacts as they were saved in the desired backup.

The Backup and Restore feature gives the administrator the ability to take “on-demand” backups of CML workspaces. Core services running in the workspace are shut down during the backup process to ensure consistency in the backup data. It is recommended that backups are taken during off-peak hours to minimize user impacts. Backing up a workspace can take tens of minutes to complete, and is dependent on how much data needs to be copied by the backup jobs.

There is currently no restriction on the number of backups one can take, and the backup snapshots are retained indefinitely in the backup service vault of the underlying cloud platform. CML workspace backup details are stored in a Backup Catalog database in the CML control plane, and these entries may be listed, viewed, deleted or restored as desired.

Restoring a backup creates a new CML workspace wherein the restored data is automatically imported. All the projects, jobs, applications, etc., that were in existence during the backup are automatically available in the new workspace. Since restoring a CML backup needs to provision a new cluster, and then launch restore jobs to create storage volumes from the backup snapshots, the restore process takes a bit longer than a regular workspace provisioning operation.

### Limitations

#### Cloud Platform

Backup and Restore feature is available on AWS only. Azure support is in the early stages of investigation.

*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.*

## CDP Environment

The technical preview release supports backing up and restoring in the same CDP Environment only.

## CML Workspace Endpoint

The restored workspace has its own network identity, and is different from that of the workspace that was backed up. Applications and ML models must therefore be accessed using the new domain name of the restored workspace.

## CLI Only User Interface

The technical preview release provides CDP CLI user interface. Graphical UI is in development for the GA release.

## Prerequisites

### CDP Entitlement

The feature is gated behind the `ML_ENABLE_BACKUP_RESTORE` entitlement. This entitlement must be granted to your CDP account before trying to backup your workspace.

### AWS Backup Service Opt-in

Login to your AWS account and navigate to the AWS Backup service console. Make sure the AWS region matches the region where you will be (or are) provisioning your CML workspace. Click on Settings in the navigation pane and ensure that, in the Service opt-in table, EBS and EFS services are enabled for protection by the AWS Backup service, like the following screenshot.

EBS	 Enabled
EC2	 Enabled
EFS	 Enabled

*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 CDP Cross-Account Role Permissions

### Backup and Restore Permission IAM Policies

#### Backup Policy

On the AWS console, navigate to the IAM service and click on Policies->Create Policy. Click on the JSON tab, and replace the default text with the following. ([Click to download](#))

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "backup:*",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "backup-storage:*",
      "Resource": "*"
    },
    {
      "Action": [
        "elasticfilesystem:DescribeFilesystems",
        "elasticfilesystem:Backup",
        "elasticfilesystem:DescribeTags"
      ],
      "Resource":
"arn:aws:elasticfilesystem:*:*:file-system/*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "ec2:DescribeSnapshots",
        "ec2:DescribeVolumes",
        "ec2:describeAvailabilityZones",
        "ec2:DescribeVpcs",
        "ec2:DescribeAccountAttributes",
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeSubnets",
        "ec2:DescribePlacementGroups",
        "ec2:DescribeInstances",
        "ec2:DescribeTags"
      ],
      "Effect": "Allow",
      "Resource": "*"
    },
  ],
}
```

*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.*

```

    "Effect": "Allow",
    "Action": [
        "ec2:CreateTags",
        "ec2>DeleteSnapshot"
    ],
    "Resource": "arn:aws:ec2:*::snapshot/*"
},
{
    "Effect": "Allow",
    "Action": [
        "ec2>DeleteSnapshot",
        "ec2>CreateSnapshot"
    ],
    "Resource": [
        "arn:aws:ec2:*::snapshot/*",
        "arn:aws:ec2:*:*:volume/*"
    ]
},
{
    "Action": [
        "ec2>DeleteSnapshot"
    ],
    "Effect": "Allow",
    "Resource": "*",
    "Condition": {
        "ForAnyValue:StringEquals": {
            "aws:CalledVia": [
                "backup.amazonaws.com"
            ]
        }
    }
},
{
    "Action": [
        "tag:GetTagKeys",
        "tag:GetTagValues",
        "tag:GetResources"
    ],
    "Effect": "Allow",
    "Resource": "*"
},
{
    "Action": [
        "iam:ListRoles",
        "iam:GetRole"
    ],
    "Effect": "Allow",
    "Resource": "*"
},
{

```

```

    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": [
      "arn:aws:iam::*:role/*"
    ],
    "Condition": {
      "StringLike": {
        "iam:PassedToService": "backup.amazonaws.com"
      }
    }
  },
  {
    "Action": [
      "kms:ListKeys",
      "kms:DescribeKey",
      "kms:GenerateDataKey",
      "kms:ListAliases"
    ],
    "Effect": "Allow",
    "Resource": "*"
  },
  {
    "Action": [
      "kms:CreateGrant"
    ],
    "Effect": "Allow",
    "Resource": "*",
    "Condition": {
      "ForAnyValue:StringEquals": {
        "kms:EncryptionContextKeys":
"aws:backup:backup-vault"
      },
      "Bool": {
        "kms:GrantIsForAWSResource": true
      },
      "StringLike": {
        "kms:ViaService": "backup.*.amazonaws.com"
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": "iam:CreateServiceLinkedRole",
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "iam:AWSServiceName": "backup.amazonaws.com"
      }
    }
  }
}

```

*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.*

```
    ]
  }
}
```

Save this policy as cml-backup-policy.

### Restore Policy

On the AWS console, navigate to the IAM service and click on Policies->Create Policy. Click on the JSON tab, and replace the default text with the following. ([Click to download](#))

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateVolume",
        "ec2>DeleteVolume"
      ],
      "Resource": [
        "arn:aws:ec2:*::snapshot/*",
        "arn:aws:ec2:*:*:volume/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*::snapshot/*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSnapshots",
        "ec2:DescribeVolumes"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "elasticfilesystem:Restore",
        "elasticfilesystem:CreateFilesystem",
        "elasticfilesystem:DescribeFilesystems",
        "elasticfilesystem>DeleteFilesystem",
        "elasticfilesystem:TagResource"
      ],
      "Resource":
```

*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.*

```

"arn:aws:elasticfilesystem:*:*:file-system/*"
  },
  {
    "Effect": "Allow",
    "Action": "kms:DescribeKey",
    "Resource": "*"
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt",
      "kms:Encrypt",
      "kms:GenerateDataKey",
      "kms:ReEncryptTo",
      "kms:ReEncryptFrom"
    ],
    "Resource": "*",
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "ec2.*.amazonaws.com",
          "elasticfilesystem.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": "kms:CreateGrant",
    "Resource": "*",
    "Condition": {
      "Bool": {
        "kms:GrantIsForAWSResource": "true"
      }
    }
  }
]
}

```

Save this as cml-restore-policy.

## Trust Relationship

AWS Backup service needs to be able to assume the AWS cross-account role that is used by the CDP control plane to manage AWS cloud resources. To enable this, add the following trust relationship to your AWS cross-account role's Trust relationships (navigate to the IAM service console, then find your cross-account role by clicking on Roles):

*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.*

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "backup.amazonaws.com"
  },
  "Action": "sts:AssumeRole"
}
```

## Attach Backup and Restore Policies to the Cross-Account Role

While still on the configuration page of your cross-account role in the IAM console, click on the Permissions tab. Click on the Attach policies button to attach the `cml-backup-policy` and `cml-restore-policy` policies created above. This step ensures that the AWS Backup service will have the necessary permissions to call the EBS and EFS services on behalf of the cross-account role to manage backups.

## How to Backup a CML Workspace

You take an on-demand backup of a CML workspace using the following CDP CLI command:

```
$ cdp ml backup-workspace --workspace-crn <workspace_crn>
--backup-name <some-name>
```

where `workspace_crn` is the CRN (Cloudera Resource Name) of the workspace, and `some-name` is any human-readable, unique name to identify the backup instance. Each backup instance of the workspace is identified by a Backup CRN, which is returned by the above command. For example

```
$ cdp ml backup-workspace --workspace-crn
crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspace
:54df20f9-d877-4da8-8a37-de5561401a02 --backup-name
my-first-backup
{
  "backupCrn":
"crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspac
e_backup:dd09a0eb-5c2e-479e-9b4c-06dd710f5c34",
  "vaultName": "mlx-54df20f9-d877-4da8-8a37-de5561401a02"
}
```

You can see that the `backup-workspace` returns two strings: the backup CRN and the name of the backup vault holding the backup snapshots or recovery points. If you login to your AWS Backup console, and click on “Backup vaults”, you will see a vault with the name returned by the above command.

*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.*

A backup vault is created for each workspace that gets backed up. Multiple backups of the same workspace are stored in the same backup vault.

## Backup Catalog Commands

Once a workspace has been backed up, details of the backup are stored in a backup catalog database in the CML control plane. Backup catalog entries can be viewed, restored, or deleted.

### List Backup Instances of a Workspace

Use the following command to list the backups for a given CML workspace:

```
$ cdp ml list-workspace-backups --workspace-crn <workspace_crn>
```

Or

```
$ cdp ml list-workspace-backups --workspace-name <workspace_name>
--environment-name <env_name>
```

For example,

```
$ cdp ml list-workspace-backups --workspace-crn
crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspace:54df
20f9-d877-4da8-8a37-de5561401a02
{
  "backups": [
    {
      "backupCrn":
"crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspace_bac
kup:dd09a0eb-5c2e-479e-9b4c-06dd710f5c34",
      "backupName": "test-applications-2",
      "createdAt": "2022-01-26T03:22:10.925000+00:00",
      "creatorCrn":
"crn:altus:iam:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:user:c96f8
c8b-916b-4626-b910-47aca00593fa",
      "workspaceVersionAtBackup": "2.0.26-b171",
      "backupStatus": "succeeded"
    }
  ]
}
```

### Delete Backup Instances of a Workspace

Use the following command to delete a CML workspace backup:

```
$ cdp ml delete-backup --backup-crn
```

*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.*

For example,

```
$ cdp ml delete-backup --backup-crn
crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspace
_backup:dd09a0eb-5c2e-479e-9b4c-06dd710f5c34
{
  "workflowId": "dff443d6-7375-46da-93d6-989cb4775516"
}
```

## How to Restore a CML Workspace

Recall that restoring a CML workspace backup provisions a new CML workspace. It is therefore necessary to specify provisioning parameters for the new workspace to which restored storage volumes will be attached.

It is simplest to generate the JSON input skeleton into a file, and pass the file to the restore command.

```
$ cdp ml restore-workspace --generate-cli-skeleton >
restore-input.json
```

Customize the JSON file, and then execute

```
$ cdp ml restore-workspace --cli-input-json file://restore-input.json
```

Here is a sample JSON configuration file. ([Click to download](#))

```
{
  "newWorkspaceParameters": {
    "environmentName": "eng-ml-int-env-aws",
    "workspaceName": "my-restored-workspace",
    "usePublicLoadBalancer": false,
    "disableTLS": false,
    "provisionK8sRequest": {
      "instanceGroups": [
        {
          "instanceType": "m5.4xlarge",
          "rootVolume": {
            "size": 128
          },
          "autoscaling": {
            "minInstances": 1,
```

*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.*

```

        "maxInstances": 5,
        "enabled": true
    }
}
],
"environmentName": "eng-ml-int-env-aws",
"tags": [
    {
        "key": "my_first_key",
        "value": "my_first_value"
    },
    {
        "key": "my_second_key",
        "value": "my_second_value"
    }
],
"network": {
    "topology": {
        "subnets": [
        ]
    }
},
"useLegacyNodeLabel": false,
"xEntitlements": [
]
},
"disableSSO": false,
"enableMonitoring": true,
"enableGovernance": false,
"loadBalancerIPWhitelists": [
],
"enableModelMetrics": true,
"whitelistAuthorizedIPRanges": false,
"authorizedIPRanges": [
],
"skipValidation": false
},
"backupCrn":
"crn:cdp:ml:us-west-1:c8dbde4b-ccce-4f8d-a581-830970ba4908:workspac
e_backup:dd09a0eb-5c2e-479e-9b4c-06dd710f5c34"
}

```

## Known Issues

### 1. DSE-19502

Due to a product bug, CML applications may be shown in the UI as stuck in the 'Starting' state. Restart the application to bring the UI and application in sync.

*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.*

2. DSE-18553

Model builds need to be rebuilt before they can be deployed in the restored workspace.