Machine Learning

ML Workspaces

Date published: 2020-07-16
Date modified: 2024-04-01

Legal Notice

Contents

Provisioning ML Workspaces

This topic describes how to provision machine learning (ML) workspaces.

Before you begin

The first user to access the ML workspace after it is created must have both the MLAdmin role and the EnvironmentAdmin account role assigned. See *Configuring User Access to CML* and *Understanding account roles and resource roles* for information about this resource role.

Procedure

1. Log in to the CDP web interface.

   On Public Cloud, log in to https://console.cdp.cloudera.com using your corporate credentials or any other credentials that you received from your CDP administrator.
2. Click ML Workspaces.
3. Click Provision Workspace.
4. Fill out the following fields.
   - Workspace Name - Give the ML workspace a name. For example, *user1_dev*. Do not use capital letters in the workspace name.
   - Select Environment - From the dropdown, select the environment where the ML workspaces must be provisioned. If you do not have any environments available to you in the dropdown, contact your CDP administrator to gain access.

     Note: You cannot choose an environment when the Environment or associated DataLake and FreeIPA is not in an available or running state.
   - Existing NFS - (Azure only) Enter the mount path from the environment creation procedure.
   - NFS Protocol version - (Azure only) Specify the protocol version to use when communicating with the existing NFS server.

5. Switch the toggle to display Advanced Settings.
   a) CPU Settings - From the dropdown, select the following:
      - Instance Type: You must select an instance type that is supported by CML, or the associated validation check will fail (See *Other Settings*, below).
      - Autoscale Range
      - Root Volume Size: If necessary, you can also change the default size of the root volume disk for the nodes in the group.
   b) GPU Settings - Click the GPU Instances toggle to enable GPUs for the cluster, and set the following:
      - Instance Type: You must select an instance type that is supported by CML, or the associated validation check will fail (See *Other Settings*, below).
      - Autoscale Range
      - Root Volume Size: If necessary, you can also change the default size of the root volume disk for the nodes in the group.

      Note: In addition to the CPU and GPU instances selected here, CML also provisions two extra CPU instance groups to run infrastructure pods for the ML workspaces, as follows:

      AWS:

      - ML infrastructure node group: m5.2xlarge, with an autoscale range of 2 to 3.
      - Platform infrastructure node group: m5.large, with an autoscale range of 2 to 4.

      Azure:

      - ML infrastructure node group: Standard_D3s_v2, with an autoscale range of 2 to 3.
      - Platform infrastructure node group: Standard_D2s_v3, with an autoscale range of 2 to 4.

      These are not configurable by users.
   c) Kubernetes Config - Upload or directly enter the Kubernetes config information.
   d) Network Settings
      - Subnets for Worker Nodes: (AWS only) Optionally select one or more subnets to use for Kubernetes worker nodes.
      - Subnets for Load Balancer: Optionally select one or more subnets to use for the Load Balancer.
      - Load Balancer Source Ranges: (Azure only) Enter a CIDR range of IP addresses allowed to access the cluster.
        - If the CML workspace is provisioned with public access, enter the allowed public IP address range.
        - If the CML workspace is provisioned with private access, enter the allowed private IP address range.

        Note: When you change the Load Balancer Source Range setting, the changes are propagated to both the deployed Load Balancer in EKS and the underlying Security Group (SG).
      - Enable Fully Private Cluster: This Preview Feature provides a simple way to create a secure cluster. Only available in AWS environments in CDP.
      - Enable Public IP Address for Load Balancer

        (AWS only) You can create a load balancer with a public IP address for the private cluster. This is useful in cases where there is no VPN between the CML VPC and the customer network. In this case, the connection is over the internet.

        Note: In this network configuration, to use `kubectl` commands in the private cluster, you need to execute the commands in a network that is peered with the cluster VPN. Enabling the public IP address for the load balancer is not sufficient to allow `kubectl` commands to work.
      - Restrict access to Kubernetes API server to authorized IP ranges

        You can specify a range of IP addresses in CIDR format that are allowed to access the Kubernetes API server. By default, the Kubernetes API services of CML workspaces are accessible to all public IP addresses (0.0.0.0/0) that have proper credentials.

To specify an address to authorize, enter an address in CIDR format (for example, 1.0.0.0/0) in API Server Authorized IP Ranges, and click the plus (+) icon. In this case, the API server is accessible by the user-provided address as well as control-plane-exit-ips over the public internet.

If the feature is enabled and no IP authorized addresses are specified, then the Kubernetes API server is only accessible by control-plane-exit-ips from the public internet.

Note: Both the Amazon EKS and Azure AKS have a quota or upper limit for the maximum number of public endpoint access CIDR ranges per cluster. See the Amazon EKS service quotas or the Azure AKS documentation for more details. When the feature is enabled, the Cloudera Control Plane exit IP addresses will be automatically added to the authorized IP ranges for accessing the Kubernetes API server for CDP operations, which will use three CIDR blocks against the per-cluster limit.

- Use hostname for a non-transparent proxy

  Enter a CIDR range allowed for non-transparent proxy server access to the cluster.

e) Production Machine Learning

- Enable Governance - Must be enabled to capture and view information about your ML projects, models, and builds from Apache Atlas for a given environment. If you do not select this option, then integration with Atlas won't work.
- Enable Model Metrics - When enabled, stores metrics in a scalable metrics store, enables you to track individual model predictions, and also track and analyze metrics using custom code.

f) Other Settings

- Enable TLS - Select this checkbox if you want the workspace to use HTTPS for web communication.
- Enable public Internet access - When enabled, the CML workspace will be available on the public Internet. When disabled, it is assumed that connectivity is achieved through a corporate VPC.
- Enable Monitoring - Administrators (users with the MLAdmin role) can use a Grafana dashboard to monitor resource usage in the provisioned workspace.
- Skip Validation - If selected, validation checks are not performed before a workspace is provisioned. Select this only if validation checks are failing incorrectly.
- Tags - Tags added to cloud infrastructure, compute, and storage resources associated with this CML workspace.

  Note that these tags are propagated to your cloud service provider account. See *Related information* for links to AWS and Azure tagging strategies.
- CML Static Subdomain - This is a custom name for the workspace endpoint, and it is also used for the URLs of models, applications, and experiments. You can create or restore a workspace to this same endpoint name, so that external references to the workspace do not have to be changed. Only one workspace with the specific subdomain endpoint name can be running at a time.

  Note: The endpoint name can have a maximum of 15 characters, using alphanumerics and hyphen or underscore only, and must start and end with an alphanumeric character.

6. Click Provision Workspace.

Results

It can take up to an hour for an ML workspace to be provisioned and installed. Once the status changes to show that the workspace has been successfully provisioned, click on the workspace name to go to the web application.

Note that the domain name for the provisioned workspace is randomly generated and cannot be changed.

What to do next

Grant users access to this ML workspace using the instructions at *Configuring User Access to CML*.

Related Information

Configuring User Access to CML

Understanding account roles and resource roles

Best Practices for Tagging AWS Resources
AWS EKS cluster endpoint access control
AWS Elastic Kubernetes Service endpoints and quotas
Create an AKS cluster with API service authorized IP ranges enabled
Use tags to organize your Azure resources and management hierarchy
Use a non-transparent proxy with Cloudera Machine Learning on AWS environments

Configuring User Access to CML

This topic describes how to grant users/groups access to an environment so that they can provision and/or list ML workspaces within that environment. The same users will also be granted Single Sign-on (SSO) access to the workspaces. In addition, if a user needs to provision, upgrade, or delete an ML workspace, they also need the account-level EnvironmentAdmin role assigned. For more information, see Understanding account roles and resource roles.

Note: This topic applies to public cloud releases.

Required Role: PowerUser

There are two CDP user roles associated with the CML service: MLAdmin and MLUser. A CDP PowerUser will need to assign these roles to users who require access to the CML service within an environment.

- MLAdmin - This role grants a CDP user/group the ability to create and delete Cloudera Machine Learning workspaces within a given CDP environment. MLAdmins will also have Site Administrator level access to all the workspaces provisioned within this environment. That is, they can run workloads, monitor, and manage all user activity on these workspaces.
- MLUser - This role grants a CDP user/group the ability to list Cloudera Machine Learning workspaces provisioned within a given CDP environment. MLUsers will also be able to run workloads on all the workspaces provisioned within this environment.

For instructions, see *Granting CDP Users Access to Cloudera Machine Learning Service*.

Related Information

Understanding account roles and resource roles

Granting CDP Users Access to Cloudera Machine Learning Workspaces

This topic describes how to grant the MLAdmin and MLUser roles to users/groups that must be allowed to provision/list and access ML workspaces within a specific environment.

Note: This topic applies only to public cloud releases.

Procedure

1. Log in to the CDP web interface.
2. For a specific environment, grant the MLAdmin and MLUser roles to users/groups that must be allowed to provision/list and access ML workspaces within that environment. In addition, if a user needs to provision, upgrade, or delete an ML workspace, they also need the account-level EnvironmentAdmin role assigned. For more information, see Understanding account roles and resource roles.
   a) Click Environments.
   b) Search for the environment and navigate to the environment's Clusters page.
   c) Expand the Actions dropdown and click Manage Access.

d) Search for the user or group that requires access to the CML service in this environment and assign one of the following roles to each user/group:

- MLAdmin - This role grants a CDP user/group the ability to create and delete Cloudera Machine Learning workspaces within a given CDP environment. MLAdmins will also have Site Administrator level access to all the workspaces provisioned within this environment. That is, they can run workloads, monitor, and manage all user activity on these workspaces.

  OR
- MLUser - This role grants a CDP user/group the ability to list Cloudera Machine Learning workspaces provisioned within a given CDP environment. MLUsers will also be able to run workloads on all the workspaces provisioned within this environment.

e) Click Update Roles.

f) If necessary, search for and assign the EnvironmentAdmin role in the same way.

Note: The first user to log in to an ML workspace must always be a Site Admin (that is, a user with the MLAdmin role assigned to them). If a user assigned the MLUser role attempts to access the workspace first, the web application will display an error.

Related Information

Understanding account roles and resource roles

Granting Remote Access to ML Workspaces

This topic shows you how to allow specific users remote access to the underlying cluster that powers an ML workspace.

About this task

Note: This topic applies to AWS public cloud. On Azure public cloud, a user with the MLAdmin role can download the `kubeconfig` file, and this file alone grants access to any user who has it.

Required Role: MLAdmin

Before you begin

As part of this process, you will be required to enter the user's Amazon Resource Name (ARN). Make sure you have access to this information before you begin. Either get the ARN from the user OR look up a user's ARN in your AWS account. For the latter, go to your organisation's AWS AccountIdentity and Access Management (IAM)Users and lookup the user. The ARN is available on their Summary page.

If you are using the AWS CLI, you can run the following command to get the ARN:

```
aws sts get-caller-identity
```

```
#Sample output
{
"UserId": "ABCDE12345FGHIJKLMNO6789",
"Account": "888888888888",
"Arn": "arn:aws:iam::888888888888:user/<username>"
}
```

Procedure

1. Log in to the CDP web interface.
2. Click ML Workspaces.

3. Click Actions to expand the dropdown menu.
4. Click Manage Remote Access.
5. Enter the user's ARN, or select the user's name.
6. Click Grant Access.

   To remove access for a user, in the Actions column, click Revoke Access next to the user's name.
7. Click Download Kubeconfig.

What to do next

Send the downloaded Kubernetes config file to the user who has been granted access. To be able to connect to the EKS cluster, they will need to have aws-iam-authenticator installed.

Related Information

Installing aws-iam-authenticator (AWS Documentation)

Accessing ML Workspaces via SOCKS Proxy

This topic describes how to configure a SOCKS proxy to access ML workspaces on non-publicly routable VPCs. A SOCKS proxy server allows your web browser to connect directly and securely to your ML workspaces without exposing their ports outside the subnet.

About this task

Note: This topic applies to public cloud releases.

Procedure

1. In the non routable VPC, create an EC2 instance for your SOCKS server (for example, *my-ec2-socks-server*) with a public IP and an SSH key-pair (for example, *my-key-file.pem*).

   Use the AWS documentation to create the EC2 instance: Getting Started with Amazon EC2 Instances

   Depending on whether you want multiple users to share the SOCKS server or have everyone create their own server, pick the SSH key pair for the instance accordingly. More information is available in the AWS documentation: Amazon EC2 Key Pairs.
2. Set up a SOCKS proxy server with SSH to access the EC2 instance, *my-ec2-socks-server*.

```
nohup ssh -i
        "my-key-file.pem" -CND 8157
        ec2-user@<public ip for my-ec2-socks-server> &
```

   - nohup (optional) is a POSIX command to ignore the HUP (hangup) signal so that the proxy process is not terminated automatically if the terminal process is later terminated.
   - *my-key-file*.pem is the private key you used to create the EC2 instance where the SOCKS server is running.
   - C sets up compression.
   - N suppresses any command execution once established.
   - D 8157 sets up the SOCKS 5 proxy on the port. (The port number 8157 in this example is arbitrary, but must match the port number you specify in your browser configuration in the next step.)
   - ec2-user is the AMI username for the EC2 instance. The AMI username can be found in the details for the instance displayed in the AWS Management Console on the Instances page under the Usage Instructions tab.
   - *<public ip for my-ec2-socks-server>* is the public IP address of the EC2 instance running the SOCKS server.
   - & (optional) causes the SSH connection to run as an operating system background process, independent of the command shell. (Without the &, you leave your terminal open while the proxy server is running and use another terminal window to issue other commands.)

3. Configure Your Browser to Use the Proxy. This example uses Google Chrome.

   By default, Google Chrome uses system-wide proxy settings on a per-profile basis. To get around that you can start Chrome using the command line and specify the following:

   - The SOCKS proxy port to use (must be the same value used in step 1)
   - The profile to use (this example creates a new profile)

   This creates a new profile and launches a new instance of Chrome that does not interfere with any currently running instance.

   - Linux

     ```
     /usr/bin/google-chrome \
     --user-data-dir="$HOME/chrome-with-proxy" \
     --proxy-server="socks5://localhost:8157"
     ```

   - MacOS

     ```
     "/Applications/Google Chrome.app/Contents/MacOS/Google Chrome" \
     --user-data-dir="$HOME/chrome-with-proxy" \
     --proxy-server="socks5://localhost:8157"
     ```

   - Windows

     ```
     "C:\Program Files (x86)\Google\Chrome\Application\chrome.exe" ^
     --user-data-dir="%USERPROFILE%\chrome-with-proxy" ^
     --proxy-server="socks5://localhost:8157"
     ```

Results

You should now be able to navigate to any ML workspace in the browser launched using SOCKS proxy.

When you connect to the ML workspace, the browser actually connects to the proxy server, which performs the required SSH tunneling.

Monitoring ML Workspaces

This topic shows you how to monitor resource usage on your ML workspaces.

About this task

Cloudera Machine Learning leverages Prometheus and Grafana to provide a dashboard that allows you to monitor how CPU, memory, storage, and other resources are being consumed by ML workspaces. Prometheus is an internal data source that is auto-populated with resource consumption data for each workspace. Grafana is a monitoring dashboard that allows you to create visualizations for resource consumption data from Prometheus.

Each ML workspace has its own Grafana dashboard.

Before you begin

Required Role: MLAdmin

You need the MLAdmin role to view the Workspace details page.

Note: On Private Cloud, the corresponding role is EnvironmentAdmin.

Procedure

1. Log in to the CDP web interface.

2. Click ML Workspaces.
3. For the workspace you want to monitor, click Actions Open Grafana .

Results

CML provides you with several default Grafana dashboards:

- K8s Cluster: Shows cluster health, deployments, and pods
- K8s Containers: Shows pod info, cpu and memory usage
- K8s Node: Shows node cpu and memory usage, disk usage and network conditions
- Models: Shows response times, requests per second, cpu and memory usage for model replicas.

You might choose to add new dashboards or create more panels for other metrics. For more information, see the *Grafana documentation*.

What to do next

Note: Prometheus captures data for the previous two weeks.

Related Information

Grafana documentation

Monitoring and Alerts

Suspend and resume ML workspaces

Cloud consumption costs are a pain point for many public cloud users. The CML Suspend feature allows users to scale down the Kubernetes pods running on CML infra and CPU/GPU nodes for a given ML workspace. When the resume operation is performed on the suspended workspace, the suspended pods scale up.

About this task

A suspended CML workspace has all its autoscaling node groups, except the Platform Infra node group, shrunk to zero instances, thereby saving compute instance costs for the duration the workspace is suspended. However, Kubernetes pods running on Platform Infra nodes continue to run when a workspace is suspended.

When a workspace is suspended, you cannot access the workspace URL, and all associated models, applications, sessions, and jobs also become unavailable. The suspend operation terminates sessions and jobs, so the suspend should be started only after those operations have finished. When the workspace is resumed, models and applications automatically resume operation at the same URLs as before.

Note: Make sure that disks are tagged to avoid garbage collection during backup, restore, upgrade, or suspend operations on CML workspaces. For more information, see *Tagging disks to avoid garbage collection*.

Procedure

1. To suspend a CML workspace, in the Workspaces UI, select Actions Suspend Workspace for the workspace to suspend. Then click OK to start the suspend process.
2. To resume a CML workspace, in the Workspaces UI, select Actions Resume Workspace for the workspace to resume. Then click OK to start the resume process.

Related Information

Tagging disks to avoid garbage collection

Suspend workspace API documentation

Backing up ML workspaces

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 .

You can backup an ML workspace, and restore it to a new workspace later. The backup preserves all files, models, applications and other assets in the workspace (files are note backed up by CML automatically for external NFS-based workspaces). All workspace backups can be viewed in the Workspace Backup Catalog UI.

The Backup and Restore feature gives you the ability to backup all of your data (except files in external NFS-backed workspaces) 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.

The time required to complete backing up a workspace depends on the amount of data to copy. The backup process copies data from both EBS volumes and EFS. In general, the time taken to backup EFS is more significant than for EBS. Due to the incremental nature of backups, the first backup always takes the longest amount of time. Subsequent backups should complete faster as they are built on top of the initial backup copy. For this reason, we recommend that CML workspaces be backed up regularly.

The time to backup EFS is highly dependent on the amount of data, and on the nature and number of files. It is also affected by available bandwidth in the AWS cloud backend. We have seen first-time backup of a 600 GB EFS file system taking around 10 hours. If you have much more than 600 GB on your EFS file system, the default backup timeout of 12 hours may not be long enough. In such cases, we recommend you take your first backup with a lower timeout, such as 2 hours. The CML Control Plane may abort the backup due to the timeout expiry. However, the Control Plane does not cancel the underlying backup jobs. You can monitor these backup jobs on the AWS Backup console, and if all eventually complete successfully, you can initiate the backup operation again from the CML Control Plane. This should complete in a relatively shorter time, and you will have a good backup copy to restore from if necessary.

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 the Workspace Backup Catalog UI 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. Restoring a CML backup provisions a new cluster, and then launches restore jobs to create storage volumes from the backup snapshots. The restore process takes longer than a regular workspace provisioning operation due to the extra work in copying data from backup to the new storage volumes. While backups are incremental, restores are always full-copy restores. The time to restore is dominated by EFS restoration, which takes at least as long as the time to backup the file system. The restored workspace is always created with the latest CML software version, which may be different from the CML version of the original workspace that was backed up.

Note: At this time, the ML workspace Backup and Restore feature is available on AWS, both through the UI and CLI. On Azure, this feature is only available through CLI.

Note: Make sure that disks are tagged to avoid garbage collection during backup, restore, upgrade, or suspend operations on CML workspaces. For more information, see *Tagging disks to avoid garbage collection*.

Note: A restored workspace from the workspace backup is considered a new separate workspace independent of the original workspace. Users' roles in the Control Plane from the original or backup workspace are not copied and assigned to the restored workspaces by the restore process to avoid security concerns and have to be moved after restore manually if intended.

Related Information

Tagging disks to avoid garbage collection

Workspace Backup and Restore Prerequisites

To backup and restore workspaces, check that the following prerequisites are satisfied.

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 have your CML workspace. Click on Settings in the navigation pane, and in the Service opt-in table, ensure that EBS and EFS services are enabled for protection by the AWS Backup service, as shown here.

EBS	Enabled
EC2	Enabled
EFS	Enabled

For additional information about this feature, see July 2022: Cloudera Customer Advisory: The new feature CML Backup and Restore on AWS requires changes to IAM permissions in their cross account roles (requires login).

AWS CDP Cross-Account Role Permissions

1. Install the Backup IAM 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 contents of the following JSON file. (Click here to download the file)

```
{
    "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": "*"
        },
        {
            "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"
                }
            }
        }
    ]
}
```

Save this policy as cml-backup-policy.

2. Install the 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 contents of the following JSON file. (Click here to download the file)

```
{
    "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": "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.

3. Set up the 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).

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

4. Attach the 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 Attach policies to attach the cml-back up-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.

Backing up an ML workspace

Backing up an ML workspace preserves all files, models, applications, and other assets in the workspace, although files in external NFS-backed workspaces are not backed up by CML automatically.

Procedure

1. In the Workspaces UI, find the workspace to back up. The workspace must be in the Installation completed state, otherwise backup is disabled.
2. Enter the workspace, and manually stop all workloads (sessions, jobs, applications, and models).

   For external NFS backed workspaces, manually back up the configured external NFS data to another location. This manual backup of the NFS data will be used when this particular backup is restored in future. Ignore this step if the workspace is configured with internal NFS, as internal NFS data is backed up and restored automatically by CML.
3. In the Actions menu for that workspace, select Backup Workspace.
4. In the Backup Workspace modal, enter a Backup Name to identify the workspace, and enter an appropriate timeout value.
5. Click Backup to start the process.

Results

The workspace shuts down, and the backup process begins. The workspace state changes to reflect the ongoing backup progress. If necessary, click Cancel to cancel the backup process. The backup process can take some time to complete, depending on the amount of data to copy.

Note: The default timeout is 12 hours. The estimated time to complete a backup (from the cloud provider) is now periodically added to the event logs.

What to do next

Monitoring event logs

You can monitor the progress of the backup process by checking the event logs. In the Actions menu for the workspace, click View Event Logs, and then on the Events & Logs tab, click View Event Logs again for the latest backup event.

When the backup process completes, the workspace enters the Installation completed state again.

If there were issues during backup, appropriate error messages will be displayed in the event logs. However the workspace will recover from failure and will be reverted back to the original state when backup was triggered.

Restore an ML workspace

Restoring a backup creates a new CML workspace, and recreates all of the projects, jobs, applications and so on in the original workspace.

About this task

Note: Restoring a workspace is a non-reversible operation. The restore process overwrites the existing workpace with older backup data. Any data in the running workspace that is not backed up will be lost. To save the current state, take a new backup before proceeding with the restore operation.

Procedure

1. In the Workspace Backups UI, find the workspace to restore. You can search for the workspace name or CRN. There can be multiple backups for a given workspace.
2. Enter the workspace, and manually stop all workloads (sessions, jobs, applications, and models).

   For external NFS backed workspaces, copy the manual backup of external NFS data (corresponding to this particular backup) to the configured external NFS export. Ignore this step if the workspace is configured with internal NFS, as internal NFS data is backed up and restored automatically by CML.
3. Look for the backup to restore, and click Restore. The restore process starts, and the workplace state changes to Creating workspace.
4. Provision a new workspace that is in the same CDP environment as the original workspace.

Results

The restore process can take some time, depending on the amount of data to copy. When it is complete, you can find the restored workspace in the Workspaces UI.

Note: If there is an issue during the restore process, the event log will show the relevant error messages. In case of error, the workspace will not recover from the failure automatically and will not revert back to the original state prior to the restore operation.

What to do next

Monitoring event logs

You can monitor the progress of the backup process by checking the event logs. In the Actions menu for the workspace, click View Event Logs, and then on the Events & Logs tab, click View Event Logs again for the latest backup event.

When the backup process completes, the workspace enters the installation completed state again.

If there were issues during backup, appropriate error messages will be displayed in the event logs. However the workspace will recover from failure and will be reverted back to the original state when backup was triggered.

Restoring to a different environment

A backup can be restored to a different CDP environment, as long as it is within the same AWS account and region. Make sure the following requirements are met:

- Environment roles must be within the same AWS account and region.
- The target environment must have the necessary restore-related permissions, entitlements, and trust relationships.
- Within the environment where the backup is stored, the user must have the ml/listWorkspaceBackups permission.
- Within the environment where the workspace will be restored, the user must have the ml/createWorkspace permission.

Removing ML Workspaces

This topic describes how to remove an existing ML workspace and clean up any cloud resources associated with the workspace. Currently, only CDP users with both the MLAdmin role and the EnvironmentAdmin account role can remove workspaces.

Procedure

1. Log in to the CDP web interface.
2. Click ML Workspaces.
3. Click on the Actions icon and select Remove Workspace.
   a) Remove EFS Storage - This option is enabled by default. If you want to retain project files on EFS, disable this property.
   b) Force Delete - This property is not required by default. You should first attempt to remove your workspace with this property disabled.

      Enabling this property will delete the workspace from CDP but does not guarantee that the underlying cloud resources used by the workspace will be cleaned up properly. Go to your cloud service provider account to make sure that the cloud resources have been successfully deleted.

      When manually cleaning up resources, make sure that the following types of shared resources are not deleted:

      AWS:

      - VPCs
      - Subnets
      - Storage (S3 buckets and bucket entries)
      - AWS IAM roles

      Microsoft Azure

      - Virtual networks
      - Subnets
      - ADLS storage
      - Azure resource groups (RGs named <liftie-id> and MC_<liftie-id>_<azure-region>)
4. Click OK to confirm.

   Note: On Azure public cloud, you also need to delete NFS storage after removing the workspace, if the NFS service is no longer needed.

Upgrading ML Workspaces

This topic describes how to upgrade existing ML workspaces. Currently, only CDP users with both the MLAdmin role and the EnvironmentAdmin account role can create, upgrade, or remove workspaces.

Existing ML workspaces periodically should be upgraded. Upgrading the workspace upgrades the CML software version to the current version, and may also upgrade cluster software. In case the underlying Kubernetes software must be upgraded, a warning banner displays, notifying you that you should upgrade the workspace promptly.

- During an upgrade, any running models and applications shut down, but they automatically restart after the upgrade is complete.
- To upgrade Kubernetes, only use the upgrade method provided in CML. Do not upgrade Kubernetes directly in the cloud console or through the CLI. Follow the instructions here to upgrade Kubernetes. If there is some error, then repeat the instructions. This applies to both Microsoft Azure and AWS.
- You should back up your workspace before starting the upgrade. For more information, see Backing up ML workspaces.

When is an upgrade necessary?

Cloud service providers define their generally available version of Kubernetes based on their Kubernetes version support policies. For AKS refer to Supported Kubernetes versions in Azure Kubernetes Service (AKS) and for EKS refer to Amazon EKS Kubernetes release calendar.

Cloud service providers may have different deprecation policies for Kubernetes versions:

- For AWS deprecation policy, refer their FAQ section in Amazon EKS version support and FAQ.
- For Azure, refer to the Azure Kubernetes FAQ.

If any Kubernetes version used in your ML workspaces is deprecated by the cloud providers and CML upgrades are enabled, the warning banner displays.

```
ACTION REQUIRED: A new CML version is available and it is highly recommended
 to upgrade
    to the latest version as soon as possible. To perform an upgrade, select
 Upgrade
     Workspace from the Actions menu.
```

In order to avoid unplanned service interruption caused by the automatic Kubernetes upgrade by EKS and continue to receive support from AKS for your ML workspaces on Azure, it is important to make sure that your ML workspaces are using supported Kubernetes versions. Upgrading a ML workspace will automatically upgrade the Kubernetes to a supported version. We recommend our users to upgrade the ML workspaces promptly when the warning banner appears.

What type of upgrades does CML Support?

In-place ML upgrades

Upgrades are done in-place on the existing CML workspace. This may involve a Kubernetes upgrade (if there is an upgrade available) followed by upgrading the CML software.

Note: Make sure to backup your workspace before starting the upgrade process. For more information, see Backing up ML workspaces.

1. Log in to the CDP web interface.
2. Click ML Workspaces.
3. For a given workspace, click on the Actions icon and select Upgrade Workspace.
4. Click OK to confirm.

The upgrade process may take anywhere between two to four hours, approximately.

Upgrades through CML Backup & Restore

If a CML workspace upgrade from a specific version could not be validated due to Kubernetes version deprecations on cloud providers or is deemed risky, in-place upgrades will be disabled for these versions.

In such cases, depending on the version of CML either the upgrade button is disabled or the in-place upgrade pre-flight check will fail, with a failure message pops up that says:

```
In-place upgrades from <existing_version> are not supported. Follow the
 documentation for the backup based upgrade steps.
```

In this case, it is recommended to go with Backup/Restore to upgrade to the latest CML version, essentially performing a workspace upgrade with all your previous data in place. Refer to ML Upgrades using Backup/Restore for more information.

Note: Make sure that disks are tagged to avoid garbage collection during backup, restore, upgrade, or suspend operations on CML workspaces. For more information, see *Tagging disks to avoid garbage collection*.

Related Information

Backing up ML workspaces
ML Upgrades using Backup/Restore
Supported Kubernetes versions in Azure Kubernetes Service (AKS)
Azure Kubernetes FAQ
Amazon EKS Kubernetes release calendar
Amazon EKS version support and FAQ
Tagging disks to avoid garbage collection

ML Upgrades using Backup/Restore

Cloudera strongly recommends following the CML release cadence by upgrading to every version soon after they are released. Following this process ensures that the CML Workspace is up to date with the latest security and bug fixes as well to benefit from new feature development. This document will take you through some considerations to be aware of before performing an upgrade, options you have when performing the upgrade, and the steps to complete the upgrade.

Before you begin

If a CML workspace upgrade from a specific version could not be validated due to Kubernetes version EOL or is deemed risky, in-place upgrades will be disabled for these versions.

In-place upgrades will be disabled from CML versions if the underlying Kubernetes versions are deprecated or going to be deprecated very soon. In such cases, depending on the version of CML either the upgrade button is disabled or the in-place upgrade pre-flight check will fail, with a failure message pops up that says: In-place upgrades from <existing_version> are not supported. Follow the documentation for the backup based upgrade steps.

In this case, it is recommended to go with ML Backup/Restore to upgrade to the latest CML version, essentially performing a workspace upgrade with all your previous data in place. Since a restore always installs the latest CML version, it essentially performs a workspace upgrade with all your existing workspace data intact. Backup/Restore is the recommended path to upgrade when a CML Workspace cannot be reliably in-place upgraded from its current version.

Prerequisites

Backup/Restore on AWS

For AWS, Backup/Restore functionality is GA, and is usable from the UI. The documentation is already available in Backing up ML workspaces. Please see the documentation for prerequisites for using Backup/Restore on AWS.

Backup/Restore on Azure

Currently, Backup/Restore in Azure is available only through the CDP CLI.

Additionally, the Backup/Restore feature does not perform a backup of NFS.

Steps to upgrade workspaces using Backup/Restore

There are five major steps to go through to upgrade older workspaces to the current version. Make sure to go through the following steps in order.

Note: Make sure that disks are tagged to avoid garbage collection during backup, restore, upgrade, or suspend operations on CML workspaces. For more information, see *Tagging disks to avoid garbage collection*.

Related Information

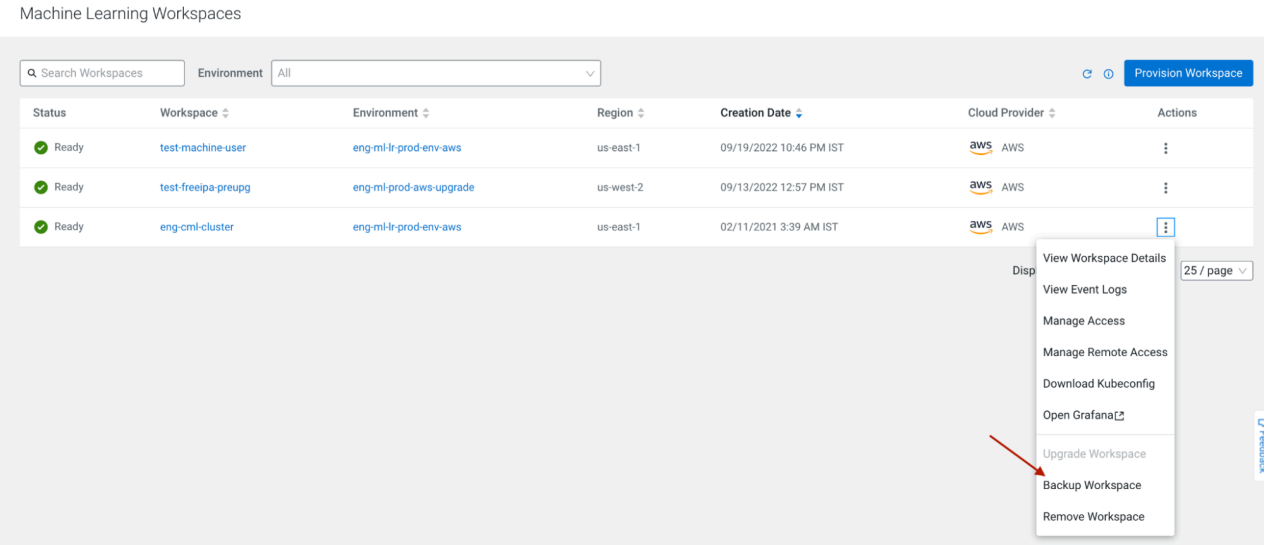
Upgrading ML Workspaces
Tagging disks to avoid garbage collection

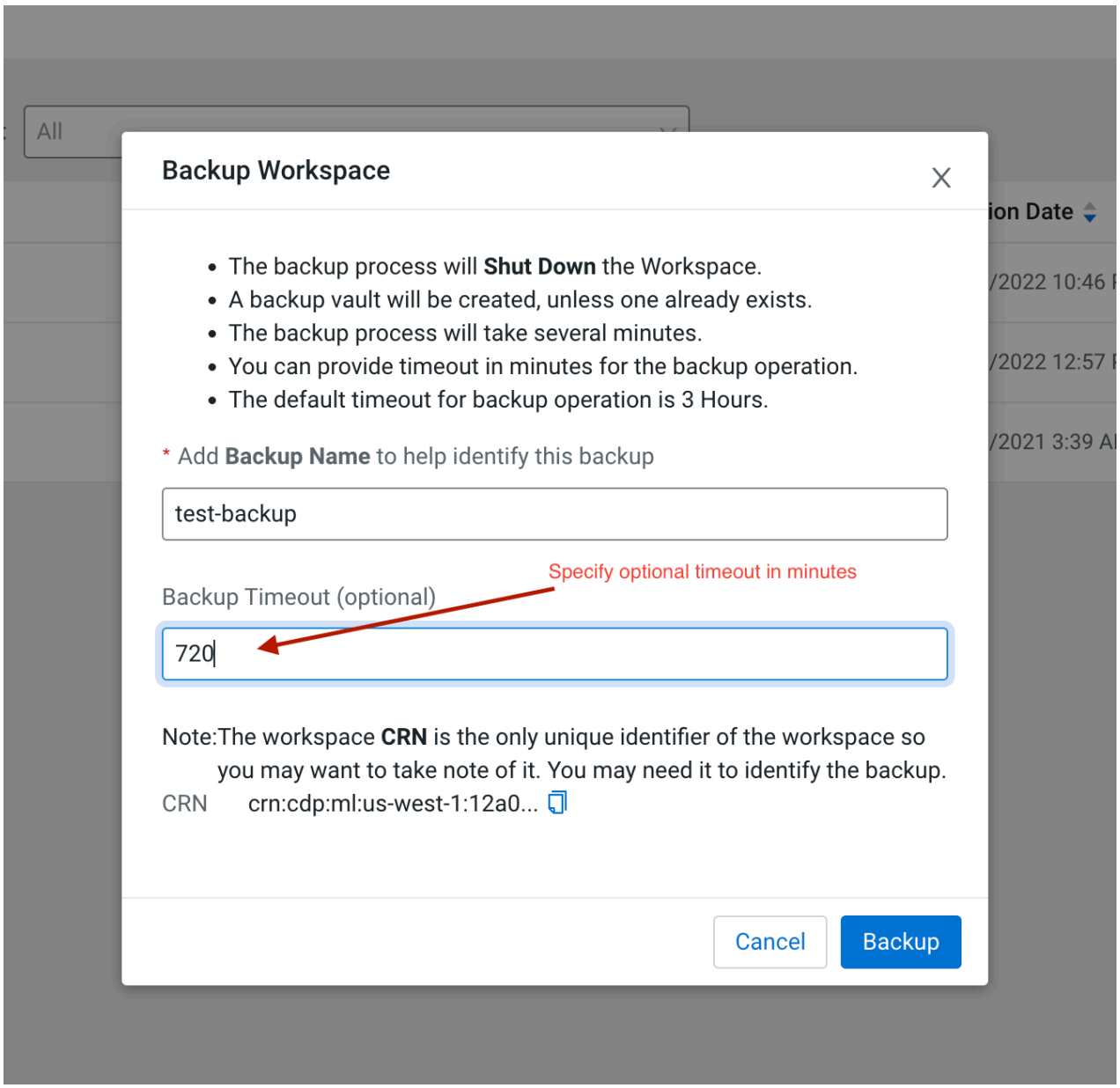
Step 1 : Backing up the workspace

After Step 1, Backing up the workspace, you should follow steps 2 through 5 in order to restore the workspace.

Backing up AWS workspace

For information on backing up workspaces, see Backing up ML workspaces.





The time required to backup or restore AWS based workspaces mainly depends on the size of EFS (File System for projects storage) associated with the workspace. To get the EFS ID associated with the workspace, click on View Workspace Details from the UI and note the Filesystem ID. The size associated with the EFS can be retrieved from the AWS console using the Filesystem ID.

The Backup timeout for storage volumes (EFS and EBS) is by default set to 12 hours, but the user can customize the timeout. While clicking on the Backup Workspace, there will be an option to specify Backup Timeout (in minutes), to accommodate users with large EFS size.

Similarly, if the user suspects that the EFS is too large to be restored within the default 12 hours, custom restore timeouts can be set in the advanced settings during restoring an AWS workspace from a backup snapshot through the UI.

Tracking running backups on AWS

For AWS based backup jobs, CML prints out backup job completion percentage for all backup jobs associated with a backup snapshot at an interval of 15 minutes as part of event logs. However, the completion percentage is an estimate returned by AWS APIs, and CML just surfaces the same. CML is not running any mechanisms/heuristics to calculate the completion percentage for a particular backup.

From our experience, we have seen that AWS-provided completion percentages can vary wildly, jump abruptly and can be downright misleading. CML advises to take the percentage numbers with a grain of salt.

Canceling long running backups on AWS

Backups generally take a long time if you have lots of data in EFS. This is expected behavior. However if the backup is taking longer than expected, you can cancel the backup jobs from AWS dashboard and CML will detect this in a while, and will fail the backup.

To cancel the corresponding AWS backup jobs from AWS dashboard:

- Go to AWS Backup Jobs Backup Jobs and identify the running backups associated with the backup snapshot. The backups should have started at approximately the same time you triggered the backup from the CML console.
- Abort all such backup jobs.

You can retry the backups again using the above mentioned steps for backing up the workspace.

Backing up Azure workspace

Prerequisites

There are a few prerequisites to Azure Backup/Restore:

- If your environment is configured with a pre-existing resource group, then CML backup service would use the same resource group for taking snapshots of Azure Disks. Else, please ensure that you have a resource group created in your Azure Account with the nomenclature `cml-snapshots-<azure_region>` . For example, if your Azure workspace resides in the `westus2` region, there should be a resource group present named `cml-snapshots-westus2`.
- Please refer to Azure documentation for roles needed to perform a backup: Use Azure role-based access control to manage Azure Backup recovery points.

Suspending the workspace

Suspend the workspace to ensure correctness of data in NFS on Azure during backup. Suspend the workspace by clicking on the Suspend Workspace option for the workspace and wait for the suspend operation to complete successfully.

Since the workspace is now in a suspended state, it is now guaranteed that no writes/mutations are happening on the NFS or Azure disks associated with the workspace.

Invoking Backup

Once you have the prerequisites sorted, please note the workspace CRN from the View Workspace Details page, and run the following command from CDPCLI to initiate workspace backup. Please replace the values in brackets with your own values.

```
$ cdp ml backup-workspace --workspace-crn <crn:cdp:ml:us-west-1:9d74eee4-1ca
d-45d7-b645-7ccf9edbb73d:workspace:792f8cfc-ba33-428d-9c80-b9bc6e799ce9> --b
ackup-name <name-of-backup-for-upgrade>

    —-
    {
    "backupCrn": ""crn:cdp:ml:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb7
3d:workspace_backup:b6cee77e-9e38-4e30-9d72-481088f43de0""
    }
```

Please note the backup CRN returned from the CLI call, as it is the backup snapshot which will be used to restore into a new workspace.

Use these variables to restore into a new workspace:

- backupCRN: The CRN returned in the response of CLI call for backup-workspace
- existingNFS: The existing NFS server path can be retrieved from View Workspace Details Filesystem ID .
- existingNFSVersion: The existing NFS version can be retrieved from View Workspace Details NFS Protocol version .

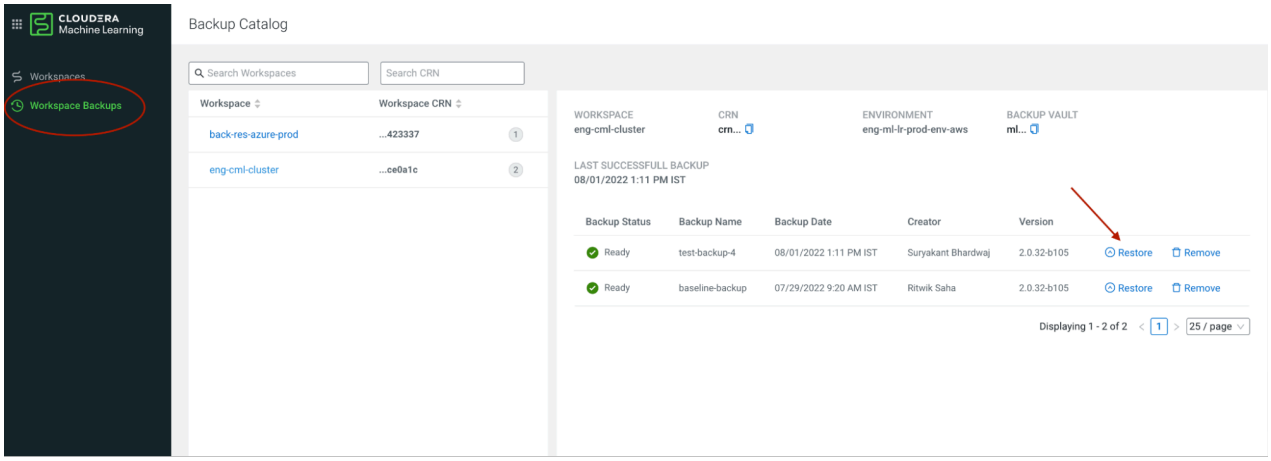
Note: Please do not perform any operations with the existing NFS in use with the suspended workspace as we will attach the same NFS in the new (restored) workspace.

Step 2: Restoring into a new workspace with a different workspace URL/ domain endpoint

Restore into a new workspace with useStaticSubdomain set to false in the CDP CLI. This brings up a workspace with a different URL/domain endpoint from the original workspace that was backed up. This step is needed to ensure that we can safely validate that restoration of the workspace is successful before executing Step 4 below to delete the original workspace. If any of the following steps fail, please contact your customer support representative.

Restore on AWS

For information on this step, see Restore an ML workspace.



The UI for restore is quite similar to the Provision workspace UI and should be familiar.

Restore on Azure

To restore into a new workspace from the backup taken above, please run the following CDP CLI command. The workspace provisioning parameters of the request needs to be configured according to your needs. Please ensure that no “write operations” are undertaken on this restored workspace since we will be using the same NFS in Step 5. This is to ensure that there is no state mismatch between the restored Azure disks and the NFS.

```
$ cdp ml restore-workspace --cli-input-json ‘{
  "newWorkspaceParameters": {
    "environmentName": "eng-ml-dev-env-azure",
    "workspaceName": "new-workspace",
    "disableTLS": false,
    "usePublicLoadBalancer": false,
    "enableMonitoring": true,
    "enableGovernance": true,
    "enableModelMetrics": true,
    "whitelistAuthorizedIPRanges": false,
    "existingNFS": "<existingNFS>",
    "nfsVersion": "<existingNFSVersion>",
    "provisionK8sRequest": {
      "instanceGroups": [
        {
          "instanceType": "Standard_DS3_v2",
          "rootVolume": {
            "size": 128
          },
          "autoscaling": {
            "minInstances": 1,
            "maxInstances": 10
          }
        }
      ],
      "environmentName": "eng-ml-dev-env-azure",
      "tags": [],
      "network": {
        "topology": {
          "subnets": []
        }
      }
    },
  },
  "backupCrn": "<backupCRN>",
  "useStaticSubdomain": false
}
’



—-
{
    "workspaceCrn": "crn:cdp:ml:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb
73d:workspace:081ee5d2-4e82-487c-9404-2537a0ab4019"
}
```

Wait for the restore operation to succeed.

After restore, please login to the newly created workspace and verify that all projects from the original workspace are available. Do not launch any sessions or applications, create new projects, or otherwise make any changes to the workspace, as that can make changes to the NFS file system that will be incompatible with what will be restored in step 4 below.

Related Information

Restore an ML workspace

Step 3: Delete the backed-up workspace

If Step 2 completes successfully, then delete the old (original) workspace.

After all is validated, and you confirm that the projects are in place, delete the original backed up workspace. To do so, from the UI, select Remove Workspace.

Step 4: Restore into a new workspace with same URL/domain endpoint as backed up workspace

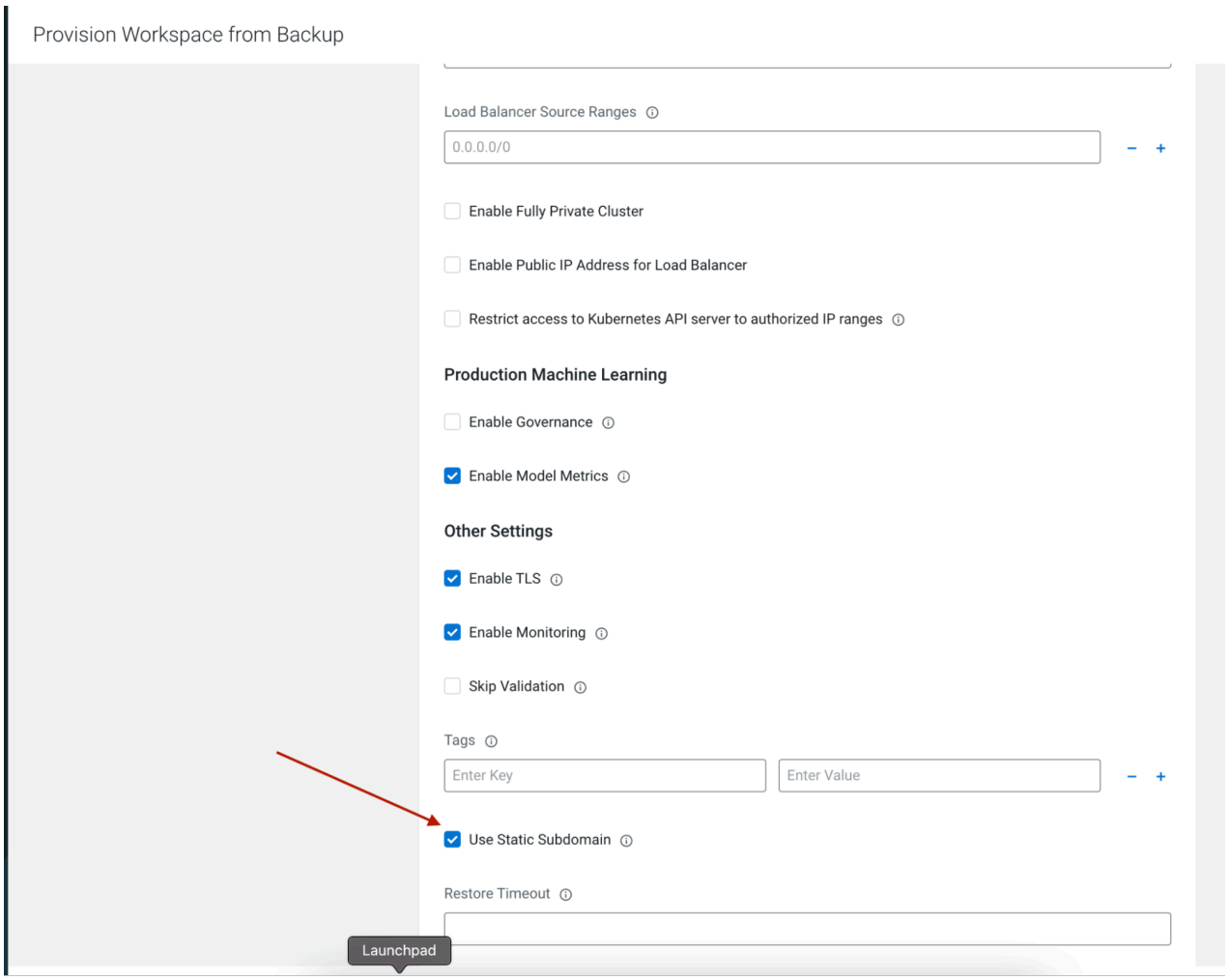
If Step 2 and 3 complete successfully, then restore into a new workspace with the same URL/endpoint as the backed-up workspace.

Since a restored workspace is a brand new workspace with data from an older workspace, a restored workspace gets a new subdomain by default. This means that any endpoints (for models, applications, etc.) that you were using from the old workspace is not valid.

To maintain the endpoints that were configured with the older workspace, check the option useStaticSubdomain in the restore payload to provision the new restored workspace with the same URL as the older one. Additionally, Use Static Subdomain is also provided as a checkbox in the Restore UI.

Restore on AWS

To restore a workspace, see Restore as ML workspace. While restoring, please ensure that Use Static Subdomain is checked in the restore UI.



Restore on Azure

To restore into a new workspace from the backup taken above, please run the following CDP CLI command. You need to tune various parameters of the request to suit workspace configuration needs.

```
$ cdp ml restore-workspace --cli-input-json '{
  "newWorkspaceParameters": {
    "environmentName": "eng-ml-dev-env-azure",
    "workspaceName": "new-workspace",
    "disableTLS": false,
    "usePublicLoadBalancer": false,
    "enableMonitoring": true,
    "enableGovernance": true,
    "enableModelMetrics": true,
    "whitelistAuthorizedIPRanges": false,
    "existingNFS": "<existingNFS>",
    "nfsVersion": "<existingNFSVersion>",
    "provisionK8sRequest": {
      "instanceGroups": [
        {
          "instanceType": "Standard_DS3_v2",
          "rootVolume": {
            "size": 128
          },
          "autoscaling": {
            "minInstances": 1,
```

```
            "maxInstances": 10
          }
        }
      ],
      "environmentName": "eng-ml-dev-env-azure",
      "tags": [],
      "network": {
        "topology": {
          "subnets": []
        }
      }
    },
  },
  "backupCrn": "<backupCRN>",
  "useStaticSubdomain": true
}
'



_-
{
    "workspaceCrn": "crn:cdp:ml:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb7
3d:workspace:081ee5d2-4e82-487c-9404-2537a0ab4019"
}
```

Related Information

Restore an ML workspace

Step 5: Delete the interim restored workspace

The upgraded workspace which was restored from the backup, in order to check the sanity of the restored workspace in Step 2, can now be safely deleted. Please identify the workspace from your control plane UI and delete the same.

Frequently Asked Questions

Some frequently asked questions about upgrading workspaces with the Backup/Restore feature.

How long does it take for a CML workspace to be backed-up and/ or restored?

CML relies on cloud provider's native services for backup/ restore. Time consumed for backup and restore depends on multiple factors such as infrastructure, network latency, data size, file structure, number of files etc and can vary across workspaces. For internal test parameters, the backup of 600GB data took approximately 10 hours on AWS.

What happens to customizations done on Kubernetes Clusters during Restore?

CML does not support applying customizations during Backup and Restore. All customizations will have to be applied through automation or manually post CML Workspace Restore.

Tagging disks to avoid garbage collection

During backup, restore, upgrade, or suspend operations on CML workspaces, EBS or Azure disks can be left in a temporarily detached or unmanaged state when the associated EKS or AKS cluster is still running. If there is a garbage collection script running, and it is not properly configured, the disks used by the workspace can be deleted unintentionally. If the disk is not backed up, then the data will be lost.

Garbage collection scripts need to check for the following tags and ignore disks that are tagged with the corresponding values.

Cloud	Tag key	Tag value
Azure	k8s-azure-created-by	kubernetes-azure-dd
AWS	kubernetes.io/cluster/liftie-*	owned

Modify Instance Group Type

You can easily change the instance type of CML workspaces, which is beneficial for optimizing performance and cost.

Key Benefits of Modifying Instance Groups:

1. Scalability and Flexibility: Scale up or down to meet user workload needs, handle peak traffic or save costs during off-peak periods.
2. Cloud Provider Compatibility: We understand that cloud providers may retire or end-of-service (EOS) certain instance types. Our Modify Instance Group feature takes this into account and allows you to seamlessly adapt to changes in the cloud provider's offerings, ensuring your CML workspace stays up-to-date.

Currently, Instance Group modification is only supported for CPU and GPU Worker instance groups.

Note: As cloud platforms do not support heterogeneous instance types within a single instance group, the current Modify Instance Group Type workflow involves deleting the existing instance group and recreating it with the desired instance type. However, it's important to note that this process may disrupt user workloads running in the user namespace of the CML Workspace, including user-created sessions, jobs, models, and applications.

Modify the Instance Group from the CDPCLI

The following example shows how to modify the instance group from the command line.

```
cdp ml modify-cluster-instance-group --workspace-crn <workspace-crn>
   --instance-group-name <instance-group-name> –instance-type <instance-typ
e>
```

Modify the Instance Group from the UI

1. Go to the Workspace Details page.
2. Navigate to the Workspace Instances section on the Workspace Details page.
3. Click on the Edit button (as shown in Figure 1).
4. Choose the appropriate Instance Type from the list of available instance types (as shown in Figure 2).
5. Click on the Save button. A confirmation box will appear (as shown in Figure 3).
6. Click on the Ok button to modify the instance group with the chosen instance type.

Figure 1: Figure 1: Edit Workspace Instance Group option in Workspace Details Page

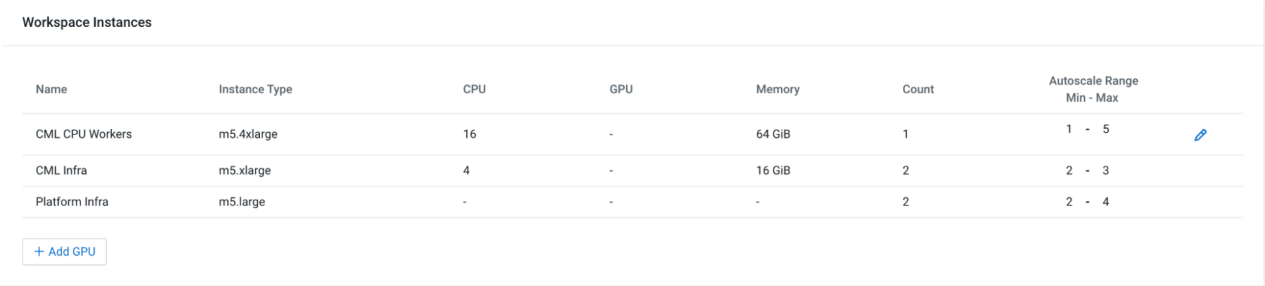


Figure 2: Figure 2: List of Instance Types for a Instance Group

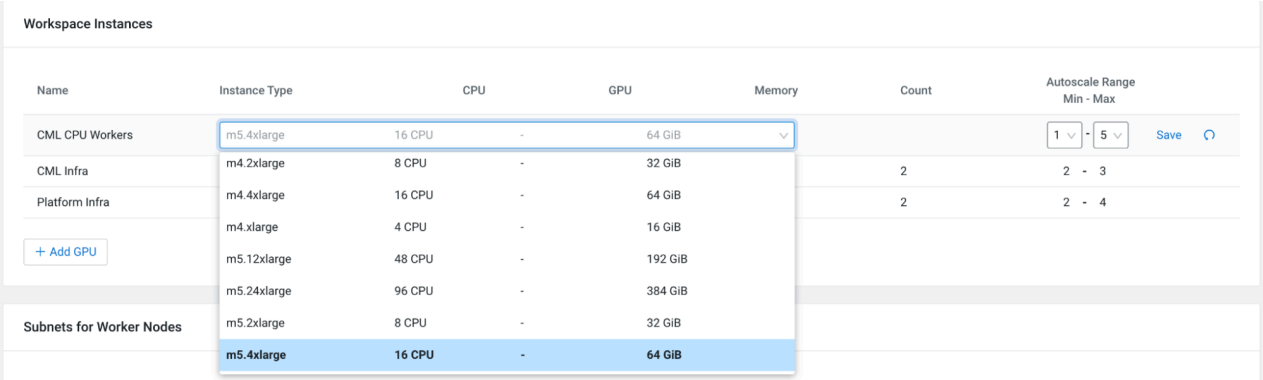
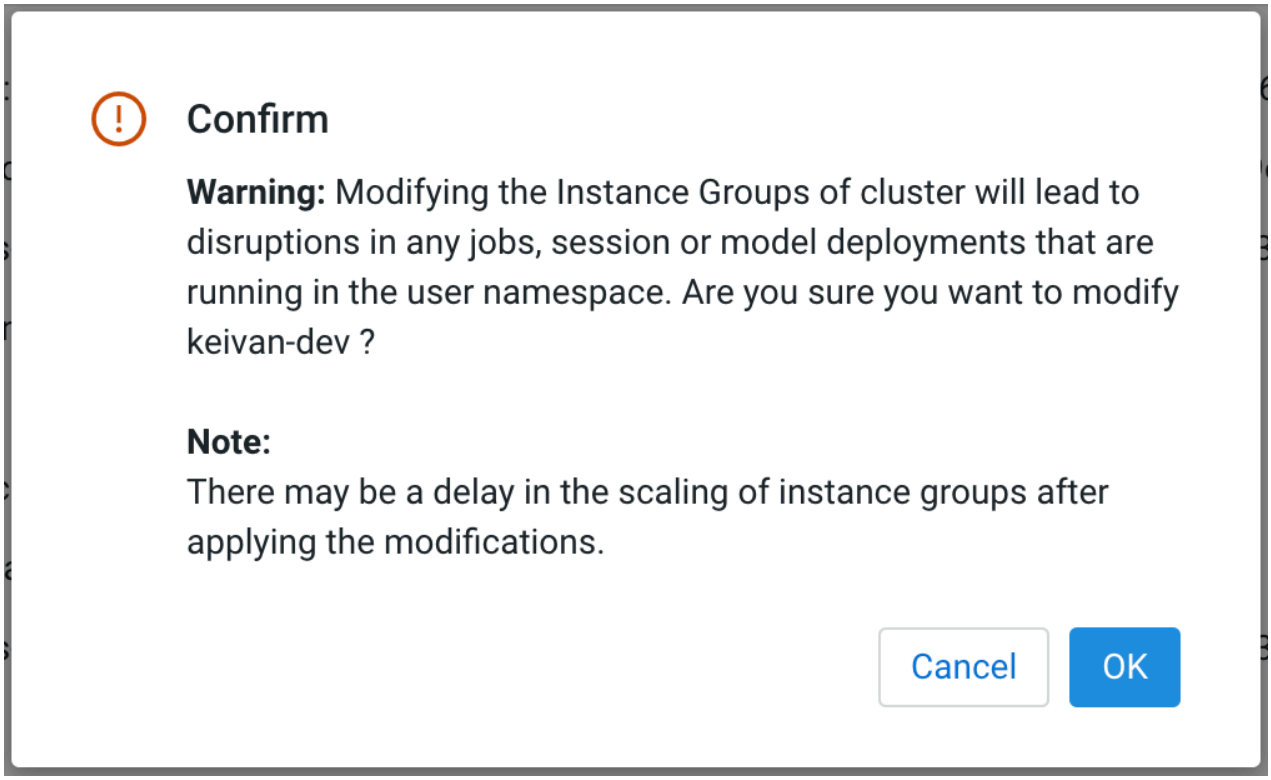


Figure 3: Figure 3: Confirmation box for Modify Instance Group Type



Note: During the modification process, there may be errors such as Bad Gateway or 404 Page not found. In this case, the applications can fail and may need to be restarted.