Cloudera Public Cloud

Azure Requirements

Date published: 2019-08-22
Date modified: 2025-08-18

CLOUDERA

Legal Notice

Contents

Azure outbound network destinations.. 83

Access to workload UIs..87

Supported browsers...88

Other resources..88

Cloudera CIDR...88

Azure subscription requirements

Before registering your Azure environment in Cloudera, use this document to verify that your Azure account has all the resources required by Cloudera and that your Cloudera administrator has adequate permissions to configure the resources and services in Azure.

As an administrator, you must be able to create and manage the resources in the Azure subscription where Cloudera users create clusters and run jobs. You must be able to perform all administrative tasks and have administrative rights to all resources. Cloudera recommends that the administrator has the role of Owner in the Azure subscription and the Application Developer role or higher in the Azure Active Directory.

Related Information

Cloudera images hosted in Azure Marketplace

Azure resources and services

Overview of Azure resources used by Cloudera

Azure outbound network access destinations

Access to workload UIs

Supported browsers

Other resources

Cloudera images hosted in Azure Marketplace

Cloudera publishes virtual hard disks (VHD) images on Azure Marketplace for each minor Runtime release. Cloudera uses these images by default during environment and Cloudera Data Hub creation on Azure cloud.

By default, Cloudera uses VHD images for deploying a Cloudera environment and Cloudera Data Hub clusters. The VHD images go through Microsoft's certification process and are then published on Azure Marketplace for each minor version of Cloudera Runtime (for example, 7.2.17). Red Hat Enterprise Linux 8 (RHEL 8) images (for Cloudera Runtime 7.2.17 and newer only) and CentOS 7 images (for Cloudera Runtime 7.2.17 and earlier only) are available.

In order for Cloudera to be able to load Cloudera-published virtual machine images in your subscription from the Azure Marketplace, you must first accept Azure Marketplace terms and conditions either via Azure CLI or Cloudera web UI. If you do not accept the terms and conditions, Cloudera cannot access the images hosted in Azure Marketplace and so it downloads them from Cloudera storage account instead.

Prior to introducing this feature, Cloudera published VHD images for Azure in regional repositories and copied these images to customer's storage account before environment creation. The images would then remain in the storage account to speed up the provisioning of future environments and clusters. If you use CentOS and you do not fulfill the requirements for Azure Marketplace images, Cloudera falls back to this storage account method.

Note: RHEL 8 images are not available using the storage account method and are only available through Azure Marketplace.

The Cloudera images hosted in Azure Marketplace allow you to:

- Deploy your Cloudera environment, Data Lake and Cloudera Data Hub clusters faster than when using an image hosted in Cloudera's storage account as there is no need to copy the VHD image anymore.
- Use RHEL images instead of CentOS images for your Cloudera environment, Data Lake and Cloudera Data Hub clusters. When using Cloudera on Azure, RHEL 8 images are only available in the Marketplace for Azure.

Cloudera recommends that for best security, performance, and cost-effectiveness you use the Azure Marketplace images.

There are two ways to accept Azure Marketplace terms and conditions:

- You can do this by enabling auto-acceptance via the Cloudera web interface. In this case, you need to do it one time only.
- You can do this via Azure CLI. In this case, you need to do it individually for each image set corresponding to a single Cloudera Runtime version.

Cloudera recommends enabling auto-acceptance via the Cloudera web interface. For more information, see Accepting the terms and conditions via Cloudera UI.

Accepting the terms and conditions via Azure CLI

You can accept Azure Marketplace terms and conditions via Azure CLI. In this case, you need to do it individually for each image offering corresponding to a single Runtime version.

Steps

1. Run the following command to obtain image URN for a specific minor Cloudera Runtime version (this is, a three-digit version such as 7.2.17):

```
az vm image list -p <NAME-OF-THE-PUBLISHER> --all -f <IMAGE-OFFER>
```

For example:

```
az vm image list -p Cloudera --all -f cdp-7_2_17
  {
    "offer": "cdp-7_2_17",
    "publisher": "cloudera",
    "sku": "runtime-7_2_17",
    "urn": "cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077",
    "version": "100.44441663.1694778077"
  },
```

2. Run the following command to print the terms and conditions details for the specific image, replacing the <SUB SCRIPTION-NAME-OR-ID> with an actual Azure subscription ID:

```
az vm image terms show \
   --urn <IMAGE-URN> \
   --subscription "<SUBSCRIPTION-NAME-OR-ID>"
```

For example:

```
az vm image terms show \
   --urn cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077 \
   --subscription "azure-eng-cloud-daily"
```

Among other information, the following details will be printed:

```
   "accepted": false,
   "id": "/subscriptions/<subscription-id>/providers/Microsoft.Marketpla
ceOrdering/offerTypes/VirtualMachine/publishers/cloudera/offers/cdp-7_2/
plans/runtime-7_2_17/agreements/current",
   "licenseTextLink": "https://mpcprodsa.blob.core.windows.net/legalterms/3
E5ED_legalterms_CLOUDERA%253a24CDP%253a2D7%253a5F2%253a24RUNTIME%253a2D7
%253a5F2%253a5F6%253a24LLJOPLSWAKNIAHKFQAUSJDXQEUIJ2TS5XNFDHNDYUPRK3LU2T
2DJPN2S32U3RVTR4DT2SVCUE5BYHA5UYFXGLSHFKCNQH7YIUS4JNXI.txt",
   "marketplaceTermsLink": "https://mpcprodsa.blob.core.windows.net/mark
etplaceterms/3EDEF_marketplaceterms_VIRTUALMACHINE%253a24AAK2OAIZEAWW5H4
MSP5KSTVB6NDKKRTUBAU23BRFTWN4YC2MQLJUB5ZEYUOUJBVF3YK34CIVPZL2HWYASPGDUY5
O2FWEGRBYOXWZE5Y.txt"
```

3. Accept the image terms by issuing the following command. The command needs to be issued individually for each image and subscription:

```
az vm image terms accept \
   --urn <image-urn> \
   --subscription "<subscription-name-or-id>"
```

For example:

```
az vm image terms accept \
   --urn cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077 \
   --subscription "azure-eng-cloud-daily"
```

Accepting the terms and conditions via Cloudera UI

You can enable auto-acceptance of Azure Marketplace terms and conditions via Cloudera web interface. This allows you to accept automatically for all Runtime versions.

Prerequisites

• You need to grant the service principal the following Azure permissions on the scope of your Azure subscription:

```
"Microsoft.MarketplaceOrdering/offertypes/publishers/offers/plans/agreemen
ts/write",
"Microsoft.MarketplaceOrdering/offerTypes/publishers/offers/plans/agreeme
nts/read"
```

> **Note:** Granting this permissions on the scope of an Azure resource group only is not sufficient. You must grant them on the scope of your Azure subscription.

• You need to grant the service principal the following Azure permission on the scope of the Cloudera Azure resource group:

```
"Microsoft.Resources/deployments/whatIf/action"
```

The Contributor role in Azure includes these permissions.

The Prerequisites for the provisioning credential: Azure permissions have been updated to include the aforementioned permissions.

Required roles

You need to be EnvironmentCreator or PowerUser in Cloudera.

Steps

Navigate to the Cloudera Management Console > Global Settings > Azure tab and check the box next to "Enable Terms & Conditions automatic acceptance". This allows Cloudera to automatically accept Cloudera image terms and conditions for Azure Marketplace:

Accepting the terms and conditions via Azure CLI

You can accept Azure Marketplace terms and conditions via Azure CLI. In this case, you need to do it individually for each image offering corresponding to a single Runtime version.

Steps

1. Run the following command to obtain image URN for a specific minor Runtime version (this is, a three-digit version such as 7.2.17):

    ```
    az vm image list -p <NAME-OF-THE-PUBLISHER> --all -f <IMAGE-OFFER>
    ```

    For example:

    ```
    az vm image list -p Cloudera --all -f cdp-7_2_17
       {
         "offer": "cdp-7_2_17",
         "publisher": "cloudera",
         "sku": "runtime-7_2_17",
         "urn": "cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077",
         "version": "100.44441663.1694778077"
       },
    ```

2. Run the following command to print the terms and conditions details for the specific image, replacing the <SUB SCRIPTION-NAME-OR-ID> with an actual Azure subscription ID:

    ```
    az vm image terms show \
       --urn <IMAGE-URN> \
       --subscription "<SUBSCRIPTION-NAME-OR-ID>"
    ```

    For example:

    ```
    az vm image terms show \
       --urn cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077 \
       --subscription "azure-eng-cloud-daily"
    ```

    Among other information, the following details will be printed:

    ```
       "accepted": false,
       "id": "/subscriptions/<subscription-id>/providers/Microsoft.Marketpla
    ceOrdering/offerTypes/VirtualMachine/publishers/cloudera/offers/cdp-7_2/
    plans/runtime-7_2_17/agreements/current",
       "licenseTextLink": "https://mpcprodsa.blob.core.windows.net/legalterms/3
    E5ED_legalterms_CLOUDERA%253a24CDP%253a2D7%253a5F2%253a24RUNTIME%253a2D7
    %253a5F2%253a5F6%253a24LLJOPLSWAKNIAHKFQAUSJDXQEUIJ2TS5XNFDHNDYUPRK3LU2T
    2DJPN2S32U3RVTR4DT2SVCUE5BYHA5UYFXGLSHFKCNQH7YIUS4JNXI.txt",
       "marketplaceTermsLink": "https://mpcprodsa.blob.core.windows.net/mark
    etplaceterms/3EDEF_marketplaceterms_VIRTUALMACHINE%253a24AAK2OAIZEAWW5H4
    MSP5KSTVB6NDKKRTUBAU23BRFTWN4YC2MQLJUB5ZEYUOUJBVF3YK34CIVPZL2HWYASPGDUY5
    O2FWEGRBYOXWZE5Y.txt"
    ```

3. Accept the image terms by issuing the following command. The command needs to be issued individually for each image and subscription:

    ```
    az vm image terms accept \
       --urn <image-urn> \
       --subscription "<subscription-name-or-id>"
    ```

    For example:

    ```
    az vm image terms accept \
       --urn cloudera:cdp-7_2_17:runtime-7_2_17:100.44441663.1694778077 \
    ```

```
--subscription "azure-eng-cloud-daily"
```

Private Azure Marketplace prerequisites

If you would like to use Private Azure Marketplace with Cloudera, there are some manual configuration steps you need to do as a prerequisite.

About this task

If you would like to use Private Azure Marketplace with Cloudera, perform the following steps outside the context of Cloudera.

Procedure

1. Assign Marketplace Admin role in the scope of Private Azure Marketplace for your admin user through the Microsoft Azure portal. For more information, see *Assign the Marketplace admin role with access control (IAM)* in the Microsoft Azure documentation linked below in the *Related information*.
2. With the same user, create ("Get") Private Azure Marketplace. For more information, see *Create private Azure Marketplace*.
3. (Optional) Create a new collection.
4. Add at least the following items to the collection:

    • Cloudera Data Platform (CDP) Freeipa
    • Cloudera Data Platform (CDP) 7.x.xx
5. Make sure the collection is enabled.
6. Make sure the collection scope contains the deployment's subscription.
7. Set Private Marketplace status to Private Marketplace enabled.
8. Launch environments, Data Lakes, or Cloudera Data Hub clusters as desired.

Results

Notifications ✕

More events in the activity log →   Dismiss all ∨

✅ **Deployment succeeded**   ✕

Deployment 'CreateVm-cloudera.freeipa-1_0-freeipa-20240614143157' to resource group 'cdp-rg' was successful.

Go to resource 📌 Pin to dashboard

2 minutes ago

✅ **Successfully updated settings**   ✕

Successfully updated Private Marketplace settings.

13 minutes ago

✅ **Successfully added items**   ✕

Successfully added 2 items:
Cloudera Data Platform (CDP) Freeipa
Cloudera Data Platform (CDP) 7.2.17

16 minutes ago

✅ **Successfully created Private Marketplace** ✕

Private Marketplace created. You can now manage your marketplace in the Manage Marketplace blade

19 minutes ago

✅ **Added Role assignment**   ✕

Peter Darvasi was added as Marketplace Admin for Microsoft.Marketplace.

20 minutes ago

For more information, see *Govern and control using private Azure Marketplace* in the Microsoft Azure documentation.

Related Information

Assign the Marketplace admin role with access control (IAM)

Create private Azure Marketplace

Govern and control using private Azure Marketplace

Azure resources and services

Cloudera uses various resources in your Azure account.

Use the following guidelines to ensure that Cloudera has access to the resources in your Azure account and that your Azure account has all the necessary resources required by Cloudera:

Prerequisites for the provisioning credential

To allow Cloudera to create resources on your Azure account, you must create app-based or certificate based credentials. The credential allows Cloudera to access and provision a set of resources in your Azure account.

Cloudera uses an app-based credential or certificate based credential to authenticate your Azure account and obtain authorization to create resources on your behalf.

The app-based credential requires that you to manually configure the service principal created within your Azure Active Directory. The app-based method requires Owner role to be able to create a service principal, which must be given Contributor role or its equivalent.

Certificate based authentication uses asymmetric authentication based on private-keys and public-keys, and does not rely on sharing secrets. When using the certificate based credential, Cloudera generates a private key and a certificate (public key) pair, and shares the certificate with you. After the certificate is shared, you must register it with Azure.

To meet Azure prerequisites for Cloudera:

1. Review the provided policies.
2. Obtain the subscription and tenant ID.
3. Create an app registration on Azure or generate and register Azure certificate credential

Azure permissions

Your Azure administrator must create custom roles in the Azure subscription.

1. The administrator must create a custom roles containing the following sets of permissions sufficient for registering an environment and creating Cloudera Data Hub clusters and Cloudera Operational Database:

   a. The administrator must create a custom role containing one of the following sets of permissions on the scope of the resource group used for Cloudera:

      1. Option 1: Use the Role definition 1: Allows Cloudera to access and use only a single existing resource group and create service endpoints on page 12 if you would like to use service endpoints.
      2. Option 2: Use the Role definition 2: Allows Cloudera to access and use only a single existing resource group and create private endpoints on page 14 if you would like to use private endpoints.

   b. In order to use the Azure Marketplace images that Cloudera publishes, the administrator also needs to grant the service principal the additional Azure permissions on the scope of your Azure subscription. The Role definition for Azure Marketplace images on page 16 policy includes these permissions.

2. Additionally, if you would like to provision other Cloudera services (Cloudera Data Engineering, Cloudera DataFlow, Cloudera Data Warehouse, or Cloudera AI), you should assign the built-in Contributor Azure role

either at the resource group level (if you are providing your own resource group) or at the Azure subscription level (if Cloudera is creating resource groups).

- If you need Cloudera Data Warehouse only and Cloudera is creating resource groups, you can use the minimal policy documented in Cloudera Data Warehouse documentation.
- If you need Cloudera DataFlow only and Cloudera is creating resource groups, you can use the minimal policy documented in Cloudera DataFlow documentation.
- If you need Cloudera AI only and Cloudera is creating resource groups, you can use the minimal policy documented in Cloudera AI documentation.

Role definition 1: Allows Cloudera to access and use only a single existing resource group and create service endpoints

The following role definition allows Cloudera to create resources only within the specified resource group:

```
{
    "Name": "Cloudera Management Console Azure Operator For Single Resource
Group",
    "IsCustom": true,
    "Description": "Can use Cloudera Management Console managed clusters and
resources updated for single resource group.",
    "Actions": [
      "Microsoft.Storage/storageAccounts/read",
      "Microsoft.Storage/storageAccounts/write",
      "Microsoft.Storage/storageAccounts/blobServices/write",
      "Microsoft.Storage/storageAccounts/blobServices/containers/delete",
      "Microsoft.Storage/storageAccounts/blobServices/containers/read",
      "Microsoft.Storage/storageAccounts/blobServices/containers/write",
      "Microsoft.Storage/storageAccounts/fileServices/write",
      "Microsoft.Storage/storageAccounts/listkeys/action",
      "Microsoft.Storage/storageAccounts/regeneratekey/action",
      "Microsoft.Storage/storageAccounts/delete",
      "Microsoft.Storage/locations/deleteVirtualNetworkOrSubnets/action",
      "Microsoft.Network/virtualNetworks/read",
      "Microsoft.Network/virtualNetworks/write",
      "Microsoft.Network/virtualNetworks/delete",
      "Microsoft.Network/virtualNetworks/subnets/read",
      "Microsoft.Network/virtualNetworks/subnets/write",
      "Microsoft.Network/virtualNetworks/subnets/delete",
      "Microsoft.Network/virtualNetworks/subnets/join/action",
      "Microsoft.Network/publicIPAddresses/read",
      "Microsoft.Network/publicIPAddresses/write",
      "Microsoft.Network/publicIPAddresses/delete",
      "Microsoft.Network/publicIPAddresses/join/action",
      "Microsoft.Network/networkInterfaces/read",
      "Microsoft.Network/networkInterfaces/write",
      "Microsoft.Network/networkInterfaces/delete",
      "Microsoft.Network/networkInterfaces/join/action",
      "Microsoft.Network/networkInterfaces/ipconfigurations/read",
      "Microsoft.Network/networkSecurityGroups/read",
      "Microsoft.Network/networkSecurityGroups/write",
      "Microsoft.Network/networkSecurityGroups/delete",
      "Microsoft.Network/networkSecurityGroups/join/action",
      "Microsoft.Network/virtualNetworks/subnets/joinViaServiceEndpoint/ac
tion",
      "Microsoft.Network/loadBalancers/delete",
      "Microsoft.Network/loadBalancers/read",
      "Microsoft.Network/loadBalancers/write",
      "Microsoft.Network/loadBalancers/backendAddressPools/join/action",
      "Microsoft.Compute/availabilitySets/read",
      "Microsoft.Compute/availabilitySets/write",
      "Microsoft.Compute/availabilitySets/delete",
```

```
      "Microsoft.Compute/disks/read",
      "Microsoft.Compute/disks/write",
      "Microsoft.Compute/disks/delete",
      "Microsoft.Compute/images/read",
      "Microsoft.Compute/images/write",
      "Microsoft.Compute/images/delete",
      "Microsoft.Compute/virtualMachines/read",
      "Microsoft.Compute/virtualMachines/write",
      "Microsoft.Compute/virtualMachines/delete",
      "Microsoft.Compute/virtualMachines/start/action",
      "Microsoft.Compute/virtualMachines/restart/action",
      "Microsoft.Compute/virtualMachines/deallocate/action",
      "Microsoft.Compute/virtualMachines/powerOff/action",
      "Microsoft.Compute/virtualMachines/vmSizes/read",
      "Microsoft.Authorization/roleAssignments/read",
      "Microsoft.Resources/subscriptions/resourceGroups/read",
      "Microsoft.Resources/deployments/read",
      "Microsoft.Resources/deployments/write",
      "Microsoft.Resources/deployments/delete",
      "Microsoft.Resources/deployments/operations/read",
      "Microsoft.Resources/deployments/operationstatuses/read",
      "Microsoft.Resources/deployments/exportTemplate/action",
      "Microsoft.Resources/subscriptions/read",
      "Microsoft.ManagedIdentity/userAssignedIdentities/read",
      "Microsoft.ManagedIdentity/userAssignedIdentities/assign/action",
      "Microsoft.DBforPostgreSQL/servers/read",
      "Microsoft.DBforPostgreSQL/servers/write",
      "Microsoft.DBforPostgreSQL/servers/delete",
      "Microsoft.DBforPostgreSQL/servers/virtualNetworkRules/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/read",
      "Microsoft.DBforPostgreSQL/flexibleServers/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/delete",
      "Microsoft.DBforPostgreSQL/flexibleServers/start/action",
      "Microsoft.DBforPostgreSQL/flexibleServers/stop/action",
      "Microsoft.DBforPostgreSQL/flexibleServers/firewallRules/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/configurations/write",
      "Microsoft.Resources/deployments/cancel/action",
      "Microsoft.Resources/deployments/whatIf/action"
    ],
    "NotActions": [],
    "DataActions": [
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/re
ad",
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/writ
e",
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/delet
e",
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/add/a
ction"
    ],
    "NotDataActions": [],
    "AssignableScopes": [
      "/subscriptions/{SUBSCRIPTION-ID}/resourceGroups/{RESOURCE-GROUP-
NAME}"
    ]
  }
```

When creating the role definition, make sure to:

- Replace the {SUBSCRIPTION-ID} with your actual subscription ID.
- Replace the {RESOURCE-GROUP-NAME} with the ID of your existing resource group

Role definition 2: Allows Cloudera to access and use only a single existing resource group and create private endpoints

The following role definition allows Cloudera to create resources only within the specified resource group:

```
{
    "Name": "Cloudera Management Console Azure Operator for Single Resource
Group",
    "IsCustom": true,
    "Description": "Can use Cloudera Management Console managed clusters and
resources, updated for use with single resource group for all resources.",
    "Actions": [
      "Microsoft.Storage/storageAccounts/read",
      "Microsoft.Storage/storageAccounts/write",
      "Microsoft.Storage/storageAccounts/blobServices/write",
      "Microsoft.Storage/storageAccounts/blobServices/containers/delete",
      "Microsoft.Storage/storageAccounts/blobServices/containers/read",
      "Microsoft.Storage/storageAccounts/blobServices/containers/write",
      "Microsoft.Storage/storageAccounts/fileServices/write",
      "Microsoft.Storage/storageAccounts/listkeys/action",
      "Microsoft.Storage/storageAccounts/regeneratekey/action",
      "Microsoft.Storage/storageAccounts/delete",
      "Microsoft.Storage/locations/deleteVirtualNetworkOrSubnets/action",
      "Microsoft.Network/virtualNetworks/read",
      "Microsoft.Network/virtualNetworks/write",
      "Microsoft.Network/virtualNetworks/delete",
      "Microsoft.Network/virtualNetworks/subnets/read",
      "Microsoft.Network/virtualNetworks/subnets/write",
      "Microsoft.Network/virtualNetworks/subnets/delete",
      "Microsoft.Network/virtualNetworks/subnets/join/action",
      "Microsoft.Network/publicIPAddresses/read",
      "Microsoft.Network/publicIPAddresses/write",
      "Microsoft.Network/publicIPAddresses/delete",
      "Microsoft.Network/publicIPAddresses/join/action",
      "Microsoft.Network/networkInterfaces/read",
      "Microsoft.Network/networkInterfaces/write",
      "Microsoft.Network/networkInterfaces/delete",
      "Microsoft.Network/networkInterfaces/join/action",
      "Microsoft.Network/networkInterfaces/ipconfigurations/read",
      "Microsoft.Network/networkSecurityGroups/read",
      "Microsoft.Network/networkSecurityGroups/write",
      "Microsoft.Network/networkSecurityGroups/delete",
      "Microsoft.Network/networkSecurityGroups/join/action",
      "Microsoft.Compute/availabilitySets/read",
      "Microsoft.Compute/availabilitySets/write",
      "Microsoft.Compute/availabilitySets/delete",
      "Microsoft.Compute/disks/read",
      "Microsoft.Compute/disks/write",
      "Microsoft.Compute/disks/delete",
      "Microsoft.Compute/images/read",
      "Microsoft.Compute/images/write",
      "Microsoft.Compute/images/delete",
      "Microsoft.Compute/virtualMachines/read",
      "Microsoft.Compute/virtualMachines/write",
      "Microsoft.Compute/virtualMachines/delete",
      "Microsoft.Compute/virtualMachines/powerOff/action",
      "Microsoft.Compute/virtualMachines/start/action",
      "Microsoft.Compute/virtualMachines/restart/action",
      "Microsoft.Compute/virtualMachines/deallocate/action",
      "Microsoft.Compute/virtualMachines/vmSizes/read",
      "Microsoft.Authorization/roleAssignments/read",
      "Microsoft.Resources/subscriptions/resourceGroups/read",
      "Microsoft.Resources/deployments/read",
```

```
      "Microsoft.Resources/deployments/write",
      "Microsoft.Resources/deployments/delete",
      "Microsoft.Resources/deployments/operations/read",
      "Microsoft.Resources/deployments/operationstatuses/read",
      "Microsoft.Resources/deployments/exportTemplate/action",
      "Microsoft.Resources/subscriptions/read",
      "Microsoft.ManagedIdentity/userAssignedIdentities/read",
      "Microsoft.ManagedIdentity/userAssignedIdentities/assign/action",
      "Microsoft.DBforPostgreSQL/servers/read",
      "Microsoft.DBforPostgreSQL/servers/write",
      "Microsoft.DBforPostgreSQL/servers/delete",
      "Microsoft.DBforPostgreSQL/flexibleServers/read",
      "Microsoft.DBforPostgreSQL/flexibleServers/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/delete",
      "Microsoft.DBforPostgreSQL/flexibleServers/start/action",
      "Microsoft.DBforPostgreSQL/flexibleServers/stop/action",
      "Microsoft.DBforPostgreSQL/flexibleServers/firewallRules/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/configurations/write",
      "Microsoft.DBforPostgreSQL/flexibleServers/privateEndpointConnectio
nsApproval/action",
      "Microsoft.Network/privateDnsZones/read",
      "Microsoft.Network/privateEndpoints/read",
      "Microsoft.Network/privateEndpoints/write",
      "Microsoft.Network/privateEndpoints/delete",
      "Microsoft.Network/privateEndpoints/privateDnsZoneGroups/read",
      "Microsoft.Network/privateEndpoints/privateDnsZoneGroups/write",
      "Microsoft.DBforPostgreSQL/servers/privateEndpointConnectionsApproval/
action",
      "Microsoft.Network/privateDnsZones/A/read",
      "Microsoft.Network/privateDnsZones/A/write",
      "Microsoft.Network/privateDnsZones/A/delete",
      "Microsoft.Network/privateDnsZones/join/action",
      "Microsoft.Network/privateDnsZones/write",
      "Microsoft.Network/privateDnsZones/delete",
      "Microsoft.Network/privateDnsZones/virtualNetworkLinks/read",
      "Microsoft.Network/privateDnsZones/virtualNetworkLinks/write",
      "Microsoft.Network/privateDnsZones/virtualNetworkLinks/delete",

      "Microsoft.Network/virtualNetworks/join/action",
      "Microsoft.Network/loadBalancers/delete",
      "Microsoft.Network/loadBalancers/read",
      "Microsoft.Network/loadBalancers/write",
      "Microsoft.Network/loadBalancers/backendAddressPools/join/action",
      "Microsoft.Resources/deployments/cancel/action",
      "Microsoft.Resources/deployments/whatIf/action"
    ],
    "NotActions": [],
    "DataActions": [
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/read"
,
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/wri
te",
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/dele
te",
      "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/add/
action"
    ],
    "NotDataActions": [],
    "AssignableScopes": [
      "/subscriptions/{SUBSCRIPTION-ID}/resourceGroups/{RESOURCE-GROUP-
NAME}"
    ]
  }
```

When creating the role definition, make sure to:

- Replace the {SUBSCRIPTION-ID} with your actual subscription ID.
- Replace the {RESOURCE-GROUP-NAME} with the name of your existing resource group

Role definition for Azure Marketplace images

In order to use the Azure Marketplace images, you need to grant the service principal additional Azure permissions on the scope of your Azure subscription. The following policy includes these permissions:

```
{
    "properties": {
        "roleName": "Cloudera Management Console Azure Operator for Azure
Marketplace",
        "description": "Can use Azure Marketplace images read and accept i
mage terms if the corresponding setting is enabled in Management Console ->
Global Settings -> Azure Settings-> Terms & Conditions auto acceptance. Scop
e must be subscription level.",
        "assignableScopes": [
            "/subscriptions/{SUBSCRIPTION-ID}"
        ],
        "permissions": [
            {
                "actions": [
                    "Microsoft.MarketplaceOrdering/offertypes/publishers/off
ers/plans/agreements/read",
                    "Microsoft.MarketplaceOrdering/offertypes/publishers/off
ers/plans/agreements/write"
                ],
                "notActions": [],
                "dataActions": [],
                "notDataActions": []
            }
        ]
    }
}
```

When creating the role definition, make sure to:

- Replace the {SUBSCRIPTION-ID} with your actual subscription ID.

Obtain subscription and tenant ID

Obtain subscription and tenant ID. You need them in order to create a provisioning credential for Azure.

These steps should be performed by someone who has the Owner built-in Azure role and the Application Developer role in Azure Active Directory.

- You can obtain both the Subscription ID and Tenant ID from Azure CLI by using the following Azure CLI command:

```
az account list|jq '.[]|{"subscriptionId": .id, "tenantId": .tenantId, "
state": .state}'
```

- You can obtain your Azure Subscription ID from your Azure Portal > Subscriptions:
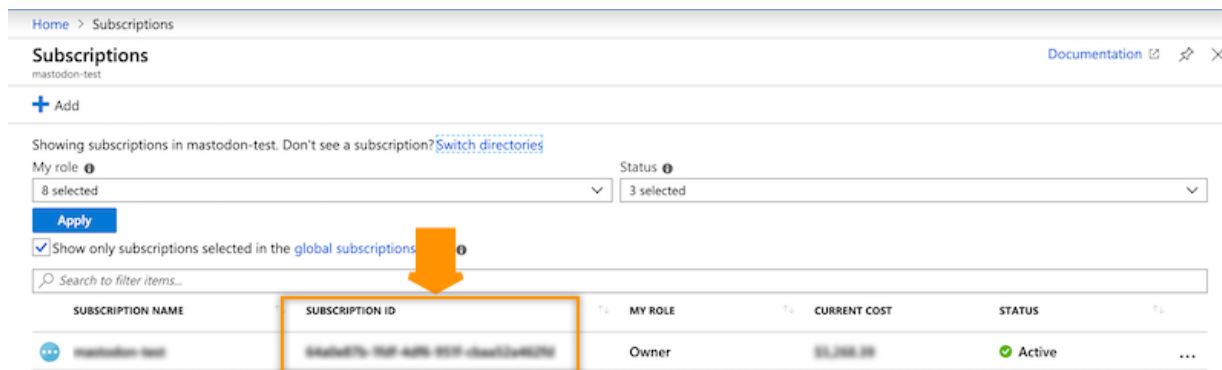
- You can obtain your Azure Tenant ID from your Azure Portal > Azure Active Directory > Properties:
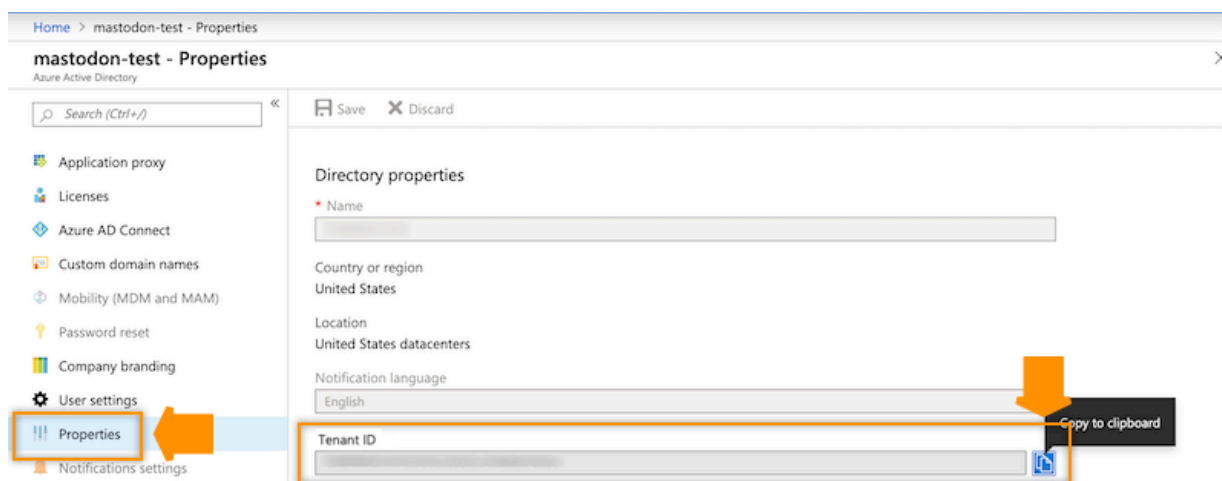
Create an app registration and assign a role to it

Create an app registration and assign a role to it. You need it in order to create a provisioning credential for Azure.

1. On Azure Portal, navigate to the Azure Active Directory > App Registrations and click on + New Registration:
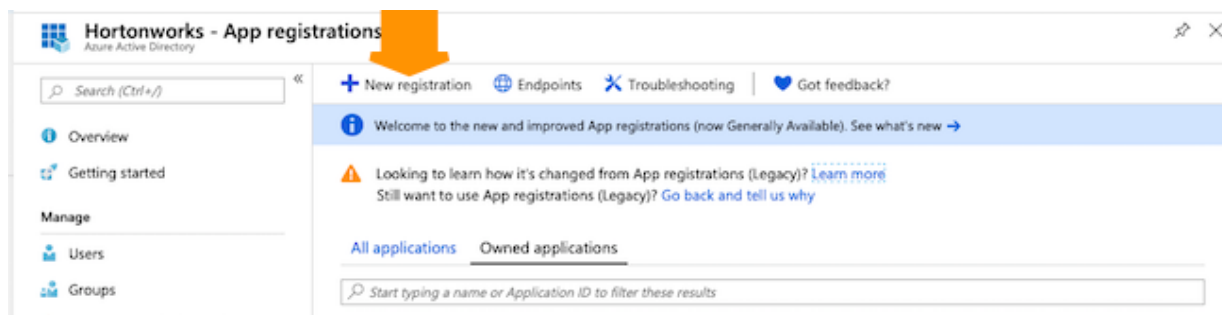
2. Register a new application as follows and then click Register:

Register an application

* Name

The user-facing display name for this application (this can be changed later).

> dominika-app ✓

Supported account types

can use this application or access this API?

- ⦿ Accounts in this organizational directory only (mastodon-test)
- ◯ Accounts in any organizational directory
- ◯ Accounts in any organizational directory and personal Microsoft accounts (e.g. Skype, Xbox, Outlook.com)

Help me choose...

Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

| Web ∨ | http://localhost ✓ |

3. Once your app registration is created, you will be redirected to the app registration's overview page. Copy and save the Application ID before closing this page. You will need to provide it to Cloudera later:

Home > App registrations > dominika-app

▦ dominika-app   📌 ✕

🔍 Search (Ctrl+/) «   🗑 Delete   ⊕ Endpoints

▦ Overview   ℹ Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? →

◢ Quickstart   Display name   Supported account types
                       dominika-app   My organization only
Manage
                       Application (client) ID   Redirect URIs
Branding   1 web, 0 public client

4. Next, navigate to Certificates & secrets and generate a new secret by clicking + New client secret, providing a description and expiration time, and clicking Add:

> **Note:** The configured expiration time for the client secret is the default expiration time of the Cloudera credentials. As environments cannot be restarted with an expired secret, ensure that the client secret is renewed before the end of the expiration time. This can be accomplished by creating a new client secret using the Azure portal or Azure CLI. After creating the new client secret, make sure to update your Cloudera credential and set the new secret.

5. Copy and save the Client secret value. You will need to provide it to Cloudera later.

6. Next, you need to assign a role to your application. To do that, browse to Subscriptions, click on your subscription, and choose Access control (IAM).

7. Click Add > Add role assignment and then assign the Contributor role or the custom role to your newly created application by:

- Under Role, selecting Contributor or the custom role.
- Typing your app name under Select and then selecting it:

8. Once done, click Save.

What to do next

Once you have this setup ready, you can Create a provisioning credential for Azure in Cloudera.

Generate and register Azure certificate credentials

You need to generate and register the certificate received by Cloudera to use certificate based authentication with Azure.

Procedure

1. Create a new Azure credential and obtain the generated certificate using the cdpcli environments create-azure-credential as shown in the following example:

```
$ cdp environments create-azure-credential --credential-name cert-example
 --app-based authenticationType=CERTIFICATE
{
    "credential": {
        "credentialName": "cert-example",
        "crn": "crn:cdp:environments:us-west-1:3D338B81-4268-48F6-B754-
C92EEE05B4B1:credential:811733C2-15A3-4228-A6E9-E2BDFC68C06A",
        "cloudPlatform": "AZURE",
        "description": "",
        "azureCredentialProperties": {
            "authenticationType": "CERTIFICATE",
            "certificate": {
                "status": "KEY_GENERATED",
                "expiration": "1732711849000",
                "expirationDate": "2024-11-27T12:50:49+00:00",
                "sha512": "<SHA512SUM STRING>",
                "base64": "<BASE64 ENCODED X509 CERTIFICATE>"
            }
        }
    }
}
```

The generated certificate is returned as a base64 string in the output.

> **Note:** The certificate has a KEY_GENERATED status. This means that it cannot be used to create Cloudera environments, Data Lakes or Data Services yet.

2. Decode the generated certificate, and save it in a file.

3. Create new app credentials in Azure using the saved certificate credentials.

   You can create the credential either in the Azure portal or you can use Azure CLI as shown in the following example:

```
$ az ad sp create-for-rbac  --name cert-creds-example --role Contributor
 --scopes /subscriptions/1B6E7D18-9A12-4800-9742-EB5451184004 --cert @cer
tificate.pem
Certificate expires 2024-11-27 12:50:49+00:00. Adjusting key credential
 end date to match.
Creating 'Contributor' role assignment under scope '/subscriptio
ns/1B6E7D18-9A12-4800-9742-EB5451184004'
The output includes credentials that you must protect. Be sure that you do
 not include these credentials in your code or check the credentials into
 your source control. For more information, see https://aka.ms/azadsp-cli
{
  "appId": "F92436F3-0E57-4EB5-BF46-18804198D726",
  "displayName": "cert-creds-example",
  "password": null,
  "tenant": "D96DC590-2EE6-42BA-B48F-1AEF42CF2BC3"
}
```
```

**4.** Update the certificate credentials with the appId, subscriptionId and tenantId to activate it as shown in the following example:

```
$ cdp environments update-azure-credential --credential-name cert-ex
ample --subscription-id 1B6E7D18-9A12-4800-9742-EB5451184004 --tenant-
id D96DC590-2EE6-42BA-B48F-1AEF42CF2BC3 --app-based authenticationType=
CERTIFICATE,applicationId=F92436F3-0E57-4EB5-BF46-18804198D726
{
    "credential": {
        "credentialName": "cert-example",
        "crn": "crn:cdp:environments:us-west-1:3D338B81-4268-48F6-B754-
C92EEE05B4B1:credential:811733C2-15A3-4228-A6E9-E2BDFC68C06A",
        "cloudPlatform": "AZURE",
        "description": "",
        "azureCredentialProperties": {
            "subscriptionId": "1B6E7D18-9A12-4800-9742-EB5451184004",
            "tenantId": "D96DC590-2EE6-42BA-B48F-1AEF42CF2BC3",
            "appId": "F92436F3-0E57-4EB5-BF46-18804198D726",
            "authenticationType": "CERTIFICATE",
            "certificate": {
                "status": "ACTIVE",
                "expiration": "1732711849000",
                "expirationDate": "2024-11-27T12:50:49+00:00",
                "sha512": "<SHA512SUM STRING>",
                "base64": "<BASE64 ENCODED X509 CERTIFICATE>"
            }
        }
    }
}
```

The certificate status is now ACTIVE. The credential can be used to create Cloudera environments, Data Lakes and Data Services.

Azure region

Prior to registering an environment, you should decide which Azure region you would like to use.

A single Azure environment registered in Cloudera corresponds to a single VNet located in a specific region, and all the resources deployed by Cloudera on Azure are deployed into that VNet.

Typically, to speed up data access, you may want to deploy clusters into the region containing the ADLS Gen2 containers that you want to access for input and output data. Therefore, when selecting the region to use, you should consider where your data is located.

Cloudera requires that the ADLS Gen2 storage location provided during environment registration is in the same region as the region selected for the environment.

If you need to use multiple regions, you need to register multiple environments, one per region.

Supported Azure regions

Cloudera supports the following Azure regions. The regions that are not mentioned below are not supported.

> **Note:** Ensure that your selected region and HA option has compute availability for Flexible Server. See Flexible Server Azure Regions.

> **Note:** Some Azure regions (such as Switzerland West, France South, Norway West, and so on) are disaster recovery regions and therefore cannot be supported by Cloudera. For more information, see Support matrix for Azure VM disaster recovery between Azure regions.

Region Name	Environment	Cloudera Data Hub	Cloudera Data Warehouse	Cloudera AI	Cloudera Data Engineering	Cloudera DataFlow	Cloudera Operational Database
Australia Central	##	##	##	##	##	##	##
Australia East	##	##	##	## (Partial GPU)	##	##	##
Australia Southeast	##	##	##	## (No ANF) (No GPU)	##	##	##
Brazil South	##	##	##	## (No ANF)	##	##	##
Canada Central	##	##	##	## (Partial GPU)	##	##	##
Canada East	##	##	##	## (No ANF) (No GPU)	##	##	##
Central India	##	##	##	## (No ANF) (Partial GPU)	##	##	##
Central US	##	##	## (AZ)	##	##	##	##
East Asia	##	##	##	## (No ANF)	##	##	##
East US	##	##	## (AZ)	##	##	##	##
East US 2	##	##	## (AZ)	##	##	##	##
Finland Central							
France Central	##	##	## (AZ)	## (No ANF)	##	##	##
Germany West Central (Public)	##	##	##	## (No ANF)	##	##	##
Israel Central	##	##					
Italy North	##	##	##	##	##	##	##
Japan East	##	##	## (AZ)	## (Partial GPU)	##	##	##
Japan West	##	##	##	## (No ANF) (No GPU)	##	##	##
Korea Central	##	##	##	## (No ANF) (Partial GPU)	##	##	##
Korea South	##	##	##	## (No ANF) (No GPU)	##	##	##
New Zealand North	##	##	##		##	##	
North Central US	##	##	##	## (No ANF)	##	##	##
North Europe	##	##	## (AZ)	##	##	##	##
Norway East	##	##	##	## (No ANF)	##	##	##
Qatar Central	##	##		##			
Poland Central	##	##	##	##	##	##	##
South Africa North	##	##	##	## (No ANF)	##	##	##
South Central US	##	##	##	##	##	##	##
South India	##	##	##	## (No ANF) (No GPU)			##
Southeast Asia	##	##	## (AZ)	##	##	##	##
Spain Central	##	##	##		##	##	
Sweden Central	##	##					

Region Name	Environment	Cloudera Data Hub	Cloudera Data Warehouse	Cloudera AI	Cloudera Data Engineering	Cloudera DataFlow	Cloudera Operational Database
Switzerland North	##	##	##	## (No ANF)	##	##	##
Switzerland West	##	##	##	##	##	##	##
UAE North	##	##	##	##	##	##	##
UK South	##	##	## (AZ)	## (Partial GPU)	##	##	##
UK West	##	##	##	## (No GPU)	##	##	##
US Gov Virginia							
West Central US	##	##			##	##	##
West Europe	##	##	## (AZ)	##	##	##	##
West India	##	##					##
West US	##	##	##	## (No ANF)	##	##	##
West US 2	##	##	## (AZ)	##	##	##	##

The regions marked with a check mark ("#")# are supported.

Note the following when reviewing regions supported for Data Warehouse:

- AZ means that a region supports availability zones.

Note the following when reviewing regions supported for Cloudera AI:

- Regions marked with a "#" and without any other annotation fulfill both ANF and GPU requirements.
- No ANF means that a region does not support Azure NetApp Files. Azure NetApp Files is required for the Cloudera AI service. If you select a region that does not include support for Azure NetApp Files, you must set up your own NFS service.
- Partial GPU means that a region supports VM types other than NCsv2. If you would like to utilize GPUs for faster computation, additional configuration will be necessary to use them.
- No GPU means that a region does not support VMs with GPUs.

Related Information

Azure geographies

Resource groups

Cloudera can provision all the environment and cluster resources into your existing resource group or you can also have Cloudera create multiple new resource groups.

Cloudera supports two resource group related scenarios:

Option	Description	Requirements	Permissions	Termination
Provide a single existing resource group	Provide a single existing resource group during environment creation and all Cloudera resources will be provisioned into that single resource group. No other resource groups will be created by Cloudera. You should select this option if you are planning to use Fine-grained access control.	If planning to use Cloudera Data Warehouse, do not use an underscore (_) when naming the resource group and use a short resource group name. The name must be fewer than 64 characters.	The scope of permissions in the role definition provided for Cloudera can be reduced to only the existing resource group where you would like Cloudera to create resources.	The resource group will not be deleted when your environment is terminated. The VHDs copied into your resource group during environment creation will not be deleted during environment termination. Cloudera preserves them to speed up subsequent environment deployments.

Option	Description	Requirements	Permissions	Termination
Cloudera creates multiple resource groups	Cloudera can create multiple resource groups. For the list of all the resource groups created, refer to Azure resources used by Cloudera. Do not use this option if you are planning to use fine-grained access control.	N/A	The scope of permissions in the role definition provided for Cloudera should be the whole subscription.	The resource groups will be deleted when your environment is terminated, except for the cloudbreak-images resource group (which stores VHDs for VM deployment).

If you would like to provide your own resource group, you can create it using the following instructions:

- Manage Azure Resource Manager resource groups by using the Azure portal
- Manage Azure Resource Manager resource groups by using the Azure CLI
- Manage Azure Resource Manager resource groups by using the Azure PowerShell

Related Information

Azure permissions

VNet and subnets

When registering an Azure environment in Cloudera, you will be asked to select a VNet and one or more subnets.

You have two options:

- Use your existing VNet and subnets for provisioning Cloudera resources.
- Have Cloudera create a new VNet and subnets. All Cloudera resources will be provisioned into this new VNet and subnets.

Note:

If you want to create an environment with cloud storage where access is allowed only from selected networks, then you must provide your own existing network. Furthermore, ensure that the network service endpoint for storage is placed on all the subnets and that the endpoint is connected to the storage account that you want to use.

Existing VNet and subnets

If you would like to use your own VNet, it needs to fulfill the following requirements:

- The VNet has at least one subnet. If you are planning to run more than Cloudera Data Hub, you need additional subnets as specified in VNet and subnet planning.
- VNet should be able to make an outbound connection with the internet or set of CIDRs and ports provided by Cloudera.
- If you would like to use private Flexible Server with delegated subnet as your Azure database server type, then you should delegate a subnet to it, as described in Private setup for Azure Flexible Server. When using Private Link, a delegated subnet is not required.
- If you would like to deploy Cloudera Data Warehouse in your environment, make sure Azure VNet subnets are large enough to support the Cloudera Data Warehouse load. When an Azure environment is activated for Cloudera Data Warehouse service, an Azure Kubernetes Service (AKS) cluster is provisioned in your subscription. The AKS cluster uses the Azure Container Networking Interface (CNI) plug-in for Kubernetes. This plug-in assigns IP addresses for every pod running inside the Kubernetes cluster. By default, the maximum number of pods per node is 30. This means that you need approximately 3,200 IP addresses for a 99-node cluster. if you activate an environment for Cloudera Data Warehouse service, make sure that the subnets are large enough on the Azure VNet for the Cloudera Data Warehouse load. Cloudera recommends using a CIDR/20 subnet or larger.
- Configure service endpoints as described in Set up service endpoint for network. If you would like to use private endpoints instead of service endpoints for Data Lake and Cloudera Data Hub, meet the requirements described in

Private endpoint for PostgreSQL; However, if you would like to use the Cloudera Data Warehouse service, you still need service endpoints regardless of what you choose for Data Lake and Cloudera Data Hub.

- If you would like to deploy Cloudera AI, note that each Cloudera AI workbench requires its own subnet.
- If you would like to deploy Cloudera Data Engineering, note that each Cloudera Data Engineering Service requires its own subnet.
- If you would like to deploy Cloudera DataFlow, note that the Cloudera DataFlow service requires its own subnet. You can have only one Cloudera DataFlow service per environment.

Warning:

Be careful if deploying into an existing subnet that has an existing Network Security Group (NSG) applied to it. In such case, there is an NSG present at VM (NIC) level and subnet level at the same time. This can cause a problem because the two settings are evaluated independently (For incoming traffic, the NSG set at the subnet level is evaluated first, then the NSG set at the VM (NIC) level. For outgoing traffic the evaluation is the reverse), and if an "allow" rule does not exist at both levels, the traffic will not be admitted. Therefore, if you need to use such a setup, you should ensure that the "allow" rule exists on both levels.

Verify the limits of the VNet and subnets available in your Azure subscription to ensure that you have enough resources to create clusters in Cloudera.

VNets can be created and managed from the Azure Portal > Virtual Networks. For detailed instructions on how to create a new VNet on Azure, refer to Create a virtual network using the Azure portal in Azure documentation.

Egress connectivity for existing VNets and subnets

When you deploy an environment with an existing network of your own configuration, it is your responsibility to create egress connectivity for the required subnets in your VNet. Egress connectivity can be accomplished through a NAT gateway setup or user-defined routing. Alternatively you can create a secondary load balancer for public egress. See Azure Load Balancers in Data Lake and Data Hub for more information.

New VNet and subnets

If you would like Cloudera to create a new VNet, you will need to specify a valid CIDR in IPv4 range that will be used to define the range of private IPs for VM instances provisioned into these subnets. Default is 10.10.0.0/16. Consider changing the IP range to correspond to corporate policies for standardized IP address ranges. The CIDR must match the <network mask>/16 pattern.

By default Cloudera creates more than 30 subnets and divides the address space as follows:

- 3 x /24 public subnets for Data Lake and Cloudera Data Hub
- 3 x /19 private subnets for Cloudera Data Warehouse
- 32 x /24 private subnets for Cloudera AI, Cloudera Data Engineering, and Cloudera DataFlow
- 3 x /19 private subnets reserved for future use

You can disable creating private subnets, in which case only 3 public subnets will be created.

If you would like to use private Flexible Server with delegated subnet as your Azure database server type in private service mode, then you should delegate a subnet to it, as described in Private setup for Azure Flexible Server. Cloudera does not create the delegated subnet for you. When using Private Link, a delegated subnet is not required.

For more information about VNet and subnets, refer to the following VNet and subnet planning documentation.

Egress connectivity for new VNets and subnets

If you are creating a new network during environment registration, Cloudera ensures that egress connectivity is available. If the "Create Public IPs" option and Public Endpoint Access Gateway are disabled in your network, a separate load balancer is created for egress, though this load balancer requires certain public IP permissions that are granted as part of the required Azure permissions. If either "Create Public IPs" or Public Endpoint Access Gateway is enabled, then a public load balancer is created to handle both public ingress to port 443 and public egress.

Azure Load Balancers in Data Lake and Cloudera Data Hub for more information.

VNet and subnet planning

Whether you decide to use your own VNet for Cloudera or have Cloudera create one for you, you should carefully plan your network, calculating and verifying the limits of the VNet and subnets available in your Azure subscription to ensure that you have enough networking resources to create clusters in Cloudera.

When registering an Azure environment in Cloudera, you are asked to select a VNet and one or more subnets. You have two options:

- Cloudera will create a new VNet and subnets
- Select a VNet and subnets that you previously created

In both cases, use this guide to calculate and verify the limits of the VNet and subnets available in your Azure subscription to ensure that you have enough networking resources to create clusters in Cloudera.

Option 1: Cloudera creates the VNet and subnets

If you would like Cloudera to create a new VNet, you will need to specify a valid CIDR in IPv4 range that will be used to define the range of private IPs for VM instances provisioned into these subnets. This must be a /16 CIDR, but you can customize the IP Range. The default is 10.10.0.0/16.

You cannot use the following reserved CIDR blocks for your VNet:

- 10.0.0.0/16
- 10.244.0.0/16
- 172.17.0.1/16
- 10.20.0.0/16
- 10.244.0.0/16

Cloudera will divide this address range as follows:

- 32 x /24 subnets - Recommended for Cloudera AI workbenches, Cloudera Data Engineering Services, and Cloudera DataFlow Services
- 3 x /19 subnets - Recommended for Cloudera Data Warehouse service
- 3 x /19 subnets - Recommended for Data Lake and Cloudera Data Hub
- 3 x /24 subnets - Reserved for future use

> **Note:**
>
> Not all subnets are used upon creation. Many of the above start getting used once you enable certain features and create more workloads in Cloudera.

If you would like to have a minimal virtual network instead, you can use the guide outlined in the next option.

Option 2: Existing VNet and subnets

If you would like to use an existing VNet, the subnet requirements vary based on the services used. Below is a guide for calculating network requirements per service.

In addition, make sure you follow the following guidelines:

- You cannot use the following reserved CIDR blocks for your VNet:

  - 10.0.0.0/16
  - 10.244.0.0/16
  - 172.17.0.1/16
  - 10.20.0.0/16
  - 10.244.0.0/16

- The Microsoft.Storage and Microsoft.SQL Service endpoints should be registered for all subnets that will be used by Cloudera.

Subnets for Data Lake and Cloudera Data Hub

Both Data Lake and Cloudera Data Hub share the same subnet, so only one subnet is required.

Cloudera recommends a minimum of a /24 CIDR. If you would like to use a smaller subnet, use the following guidelines:

- One IP address is used for each VM
- One Light Duty Data Lake cluster uses three VMs
- A typical Cloudera Data Hub cluster uses a minimum of four VMs as a starting point, but this number can be dynamically scaled up or down
- Make sure you allocate enough IPs to handle each cluster running at peak capacity

Subnets for Cloudera Data Warehouse

The Cloudera Data Warehouse service needs one subnet. You can choose the specific subnet used by Cloudera Data Warehouse when you activate Cloudera Data Warehouse for an environment. This subnet should not be shared with any of the other Cloudera applications.

Cloudera recommends a /20 or larger subnet as it can be difficult to accurately predict the size of each VW due to autoscaling.

If you would like to size the subnets to a smaller CIDR, the following guidelines assume that you are activating your Cloudera Data Warehouse environment with the default settings (no overlay networks):

VM Purpose	# VMs	# pods per VM	IPs per VM (1 for the instance + 1 per pod)	Total IP addresses required
Cloudera Data Warehouse Shared Services - (Shared among all VWs in an environment)	3	30	31	93
Per Database Catalog (One catalog is created by default, you can create additional catalogs)	2	30	31	62
Per Virtual Warehouse (XS) - without autoscaling*	2	10	11	22
Per Virtual Warehouse (S) - without autoscaling*	10	10	11	110
Per Virtual Warehouse (M) - without autoscaling*	20	10	11	220
Per Virtual Warehouse (L) - without autoscaling*	40	10	11	440

If you would like to use a smaller subnet, you can enable the Overlay Networks feature for Cloudera Data Warehouse, which will provision your AKS cluster with the kubenet network plugin instead of Azure CNI. When you use kubenet, you may need to manage the Azure Route Table (UDR) manually in accordance with AKS documentation. In this case, use the following table as your guideline:

VM Purpose	# VMs	Total IP addresses required
Cloudera Data Warehouse Shared Services (shared among all VWs in an environment)	3	3
Per Database Catalog (One catalog is created by default, you can create additional catalogs)	2	2
Per Virtual Warehouse (XS) - without autoscaling*	2	2

Per Virtual Warehouse (S) - without autoscaling*	10	10
Per Virtual Warehouse (M) - without autoscaling*	20	20
Virtual Warehouse (L) - without autoscaling*	40	40

* Each autoscaling activity can be treated as deploying a new Virtual Warehouse. For example, when a XS Virtual Warehouse is scaled once, it uses four VMs instead of two.

Subnets for Cloudera AI

Azure Files NFS v4.1 is a managed, POSIX compliant NFS service on Azure. The file share is used to store files for the Cloudera AI infrastructure and Cloudera AI workbenches. This is the recommended NFS service for use with Cloudera AI. You need one separate subnet delegated to the Azure Files NFS service (all workbenches in a region will share this service). Cloudera recommends a /28 subnet for this purpose.

You also need one subnet per each workbench that you plan to run. Cloudera AI uses the Kubenet plugin to AKS and the subnet used for a workbench cannot be shared with another workbench or another service. You can choose which subnet is used by a workbench at the time of provisioning. Cloudera recommends a /25 CIDR for these subnet, but if you would like to provide a custom range, the formula to calculate IP Addresses per workbench is as follows:

• Each workbench can grow up to 30 CPU worker nodes and 30 GPU workers; each node consumes one IP address.
• In addition, you need to allocate up to 11 IP address (6 infrastructure nodes and 5 for auxiliary networking usage).

For more information, see Network Planning for Cloudera AI on Azure.

Subnets for Cloudera Data Engineering

Cloudera Data Engineering runs in the VNet registered in Cloudera as part of your Azure environment.

Each Cloudera Data Engineering service requires its own subnet. CDE on AKS uses the Kubenet CNI plugin provided by Azure. In order to use Kubenet CNI, we need to create multiple smaller subnets when creating an Azure environment. It is recommended to partition the vnet with subnets that is just the right size to fit the expected max nodes in the cluster.

Cloudera recommends a /24 CIDR for these subnets, but if you would like to provide a custom range, the formula to calculate IP Addresses per Cloudera Data Engineering service is as follows:

• Each Cloudera Data Engineering service can scale up to 100 compute nodes; each node consumes one IP address.
• In addition, you need to allocate 3 IPs for the base infra nodes and 2 IP addresses per virtual cluster for the virtual cluster service nodes.

Subnets for Cloudera DataFlow

Cloudera DataFlow runs in the VNet registered in Cloudera as part of your Azure environment.

The Cloudera DataFlow service requires its own subnet. Cloudera DataFlow on AKS uses the Kubenet CNI plugin provided by Azure. In order to use Kubenet CNI, create multiple smaller subnets when creating an Azure environment.

Cloudera recommends a /24 CIDR for these subnets, but if you would like to provide a custom range, the formula to calculate the IP Addresses is as follows:

• Each Cloudera DataFlow service can scale up to 50 compute nodes; each node consumes one IP address.
• In addition, allocate two IPs for the base infra nodes.

Private setup for Azure Flexible Server

When Cloudera creates an Azure Database for PostgreSQL - Flexible Server instance, you must choose one of the following networking options: Private access (VNet integration or Private Link) or Public access (allowed IP addresses). Public access is used by default.

For more general information, see Networking overview for Azure Database for PostgreSQL - Flexible Server with private access (VNET Integration) and Networking overview for Azure Database for PostgreSQL - Flexible Server with public access (allowed IP addresses).

Delegated subnet

You have two options to deploy Flexible Server instances in private service mode (without public endpoints): using Private Link or without Private Link. Flexible Server with Private Link is the recommended option. If Private Link is not used, Flexible Server instances need to be deployed in a "delegated subnet" within the VNet. The Flexible Server with delegated subnet option, although still available on the UI, is deprecated and it is recommended to use Flexible Server with Private Link instead.

Delegated subnets are no longer a requirement as you can use Azure Private Link instead. However, when private endpoints are not used, it is a prerequisite to delegate a subnet to Microsoft.DBforPostgreSQL/flexibleServers as mentioned in Azure documentation. This delegation means that only Azure Database for PostgreSQL Flexible Servers can use that subnet. No other Azure resource types can be in the delegated subnet.

You need to create such a delegated subnet and provide it to Cloudera during environment registration. This delegated subnet will be used by Azure Database for PostgreSQL instances. The delegated subnet provided during environment registration will be used by default for all Azure Database for PostgreSQL instances used in Cloudera.

Note: Although you can currently select multiple subnets, the larger subnet is always used by Cloudera. That is, if there are two subnets provided, one with 128 available IPs and another with 256 available IPs, the second one will be picked. Even though you can select multiple delegated subnets, we recommend that you provide only one.

Creating a delegated subnet

For a step-by-step official guide on how to perform the delegation, see Delegate a subnet to an Azure service. For considerations on the delegated subnet's sizing, see Virtual network concepts.

Here is a screenshot from Azure Portal showing the desired setting:

The Microsoft.Storage service endpoint is set automatically during deployment by Azure.

After the subnet has successfully been delegated, don't forget to record the full subnet ID or the name of the subnet for later use as an input (referred to as <delegated-subnet-id>). For example: /subscriptions/3ddda1c7-d1f5-4e7b-ac81-abcdefgh/resourceGroups/rg/providers/Microsoft.Network/virtualNetworks/vnet/subnets/subnet

For background information, see Using Azure Database for PostgreSQL Flexible Server.

Private DNS options

Cloudera supports using VNet integration based private setup with an existing Private DNS zone that can be either pre-created and provided by you, or created by Cloudera.

When using a private setup for Azure Postgres, an Azure private DNS zone is used for the DNS service resolving the FQDN to the private IP. Cloudera offers two options. The DNS zone can be:

- Created and managed entirely by Cloudera (You select "Create new private DNS zone" during environment registration), or
- Created by you before registering an environment (You pre-create the DNS zone and select it during environment registration). You need to provide access for discovery, validation, and adding and removing DNS A records.

Private setup (with either of the two DNS options) can only be used with a single customer-provided resource group. They cannot be used with Cloudera-created multiple resource groups.

Depending on whether you prefer to bring your own DNS or have Cloudera create and manage it, refer to the following documentation:

Using Cloudera-managed private DNS

Review this documentation if you are planning to use a private setup for Azure Postgres with a Cloudera-managed DNS.

Requirements and limitations

The following limitations apply when using a Cloudera-managed private DNS:

- Only Azure's Private DNS Zone is supported. Using an on-premise DNS is not supported.
- The Private DNS Zone will be residing in the single existing resource group, even if the VNet is located elsewhere.
- Only one resource group can have private setup with a given VNet. This is because:
    - Only one DNS zone with a given name can be linked to a VNet.
    - That DNS zone is deployed in the single resource group where all the resources are located.
- The private DNS zone and virtual network links are shared within the single resource group. The first environment ever created in that resource group will create them. They will never be deleted by Cloudera.

Prerequisites

In order to use a Cloudera-managed private DNS, you should meet the following prerequisites:

1. Review DNS zones existing in your resource group on page 30
2. Ensure that Cloudera has adequate permissions on page 30

Review DNS zones existing in your resource group

If you would like Cloudera to create and manage the Private DNS Zone, review the DNS zones that exist in the resource group that you are planning to use for Cloudera and make sure that one of the following is true:

- No Private DNS Zone named privatelink.postgres.database.azure.com or flexible.postgres.database.azure.com (for Flexible Server with Private Link or Flexible Server with delegated subnet respectively) is connected to the VNnet.
- If there is a Private DNS Zone named privatelink.postgres.database.azure.com or flexible.postgres.database.azure .com (for Flexible Server with Private Link or Flexible Server with delegated subnet respectively) connected to the VNet, verify that the zone is located in the existing resource group that you are planning to use for Cloudera. If the Private DNS Zone is already used for one environment, Cloudera can reuse it for another environment.

Ensure that Cloudera has adequate permissions

Ensure that the role that you are using for the Azure credential has the permissions mentioned in Role definition 2: Allows Cloudera to use only a single existing resource group create private endpoints.

Bringing your own private DNS

Review this documentation if you are planning to use a private setup for Azure Postgres with your own private DNS.

Requirements and limitations

The following limitations apply when using your own private DNS:

- Only Azure's private DNS zone is supported. Using an on-prem DNS is not supported.
- The private DNS zone provided to Cloudera must have the name privatelink.postgres.database.azure.com in case of Flexible Server with Private Link and the name of the private DNS zone must ending with postgres.database.azure.com in case of Flexible Server with Delegated Subnet.
- The private DNS zone must be in a subscription that is accessible to the service principal used by the Cloudera app-based credential; that is, the subscription where the private DNS zone is created must be in the same tenant where the service principal is located.

Prerequisites

When bringing your own private DNS, you should meet the following prerequisites:

Create a private DNS zone and link it to your VNet

If you choose to provide your own Azure private DNS zone then you should:

1. Create a private DNS zone with a name ending with the name privatelink.postgres.database.azure.com in case of Flexible Server with Private Link or with a name ending with postgres.database.azure.com in case of Flexible Server with delegated subnet in any subscription accessible to the service principal used by the Cloudera app-based credential; that is, the subscription where the private DNS zone is created must be in the same tenant where the service principal is located.

   - In case of Flexible Server with Private Link, use the following Azure CLI command to create a private DNS zone:

   ```
   az network private-dns zone create \
     --name privatelink.postgres.database.azure.com \
     --resource-group <YOUR_DNS_RESOURCE_GROUP> \
     --subscription <YOUR_SUBSCRIPTION_FOR_DNS>
   ```

   - In case of Flexible Server with delegated subnet, use the following Azure CLI command to create a private DNS zone:

   ```
   az network private-dns zone create \
     --name flexible.postgres.database.azure.com \
     --resource-group <YOUR_DNS_RESOURCE_GROUP> \
     --subscription <YOUR_SUBSCRIPTION_FOR_DNS>
   ```

2. Link it to the VNet that you are planning to use for the Cloudera environment. You can do this using the following Azure CLI command:

   ```
   az network private-dns link vnet create \
     --name <DESIRED_LINK_NAME> \
     --resource-group <YOUR_VNET_RESOURCE_GROUP> \
     --zone-name <PRIVATE_DNS_ZONE_NAME> \
     --virtual-network <YOUR_VNET_RESOURCE_ID> \
     --subscription <YOUR_VNET_SUBSCRIPTION>
   ```

Ensure that Cloudera has adequate permissions

Ensure that the role that you are using for the Azure credential has the permissions mentioned in Role definition 2: Allows Cloudera to use only a single existing resource group and create private endpoints.

Additionally, if your DNS zone is outside of the single resource group that you are planning to provide to Cloudera, the following additional permissions are required for the service principal to be able to discover, validate, and use the

DNS zone (add and remove A records). These additional permissions need to be included in a separate role definition and assigned separately:

```
{
  "Name": "Cloudera Management Console Azure Operator for Using Private DNS
  Zones",
  "IsCustom": true,
  "Description": "Can list, validate and use private DNS zones",
  "Actions": [
  "Microsoft.Network/privateDnsZones/join/action",
  "Microsoft.Network/privateDnsZones/read",
  "Microsoft.Network/privateDnsZones/virtualNetworkLinks/read"
  ],
  "NotActions": [],
  "DataActions": [
  ],
  "NotDataActions": [],
  "AssignableScopes": [
  "/subscriptions/{SUBSCRIPTION-ID}/resourcegroups/{RESOURCE-GROUP-NAME}"
  ]
  }
```

The placeholders in bold must be replaced with actual values:

- {SUBSCRIPTION-ID} - The ID of the subscription where the DNS zone is located.
- {RESOURCE-GROUP-NAME} - The name of the resource group where the DNS zone is located.

Related Information
Quickstart: Create an Azure private DNS zone using the Azure portal

Resources created under the hood

Creating a Cloudera environment on Azure with a private Flexible Server setup created by Cloudera involves creating several Azure resources, all of which are necessary to have a private setup that works out-of-the-box.

This private setup ensures that communication between Cloudera and the Azure Postgres server happens via a private IP address. However, cluster services do need to contact the Postgres server via FQDN, so the address needs to be resolvable from the VNet.

Resources created when using a Cloudera-managed Private DNS Zone

In the scenario where Cloudera creates the private DNS, several resources are created:

- A Private DNS Zone: It is a DNS zone, part of an Azure-hosted DNS server. It has a fixed name, privatelink. postgres.database.azure.com for Flexible Server with Private Link and flexible.postgres.database.azure.com for Flexible Server with delegated subnet. However, any DNS Zone name ending with postgres.database.azure.com is usable for Flexible Server with delegated subnet in the "Bring your own private DNS" setup.
- A virtual network link between the zone and a VNet where the domain resolution should happen.
- An A record within the zone: FQDN to IP address resolution. There is no reverse lookup.

> **Note:** If there is a private DNS zone named privatelink.postgres.database.azure.com or flexible.postgres.da tabase.azure.com (for Flexible Server with Private Link or Flexible Server with delegated subnet respectively) containing the required virtual network link in the deployment's resource group, Cloudera reuses it. The A record and the network link are created only if they do not already exist.

When you delete the environment, the DNS zone and the network link that you provided will not be deleted.

Resources created when using your own private DNS zone

In this case the Private DNS Zone and the network link are provided by you and Cloudera creates the following:

- Cloudera creates an A record within the zone: FQDN to IP address resolution. There is no reverse lookup.

When you delete the environment, the DNS zone and the network link that you provided will not be deleted.

Private setup for Azure Single Server

By default, Cloudera creates and uses Azure Database for PostgreSQL - Flexible Server instance and so it requires the private setup described in Private setup for Azure Flexible Server. You should only review the documentation linked below if you specifically would like to use Single Server instead of Flexible Server; Otherwise, you can ignore this content.

When Cloudera creates an Azure Database for PostgreSQL - Single Server instance, you must choose one of the following networking options: service endpoint or private endpoint.

Service endpoint for Azure Postgres

On Azure, the external PostgreSQL database can reach the network via a service endpoint or a private endpoint. If you would like to use a service endpoint, before you can set up an external database for a given network, you must enable service endpoints for all subnets where the database should be reachable.

Note: Service endpoints are firewall rules that allow traffic only from those subnets where you explicitly granted permission. Therefore, when using service endpoints for Azure Postgres, the external database still needs to have public access enabled.

Set up a service endpoint via Azure Portal

1. From the Azure portal, go to the VNet for which you want to add service endpoints.
2. From the menu in the left pane, select Service endpoints.
3. In the Service endpoints window, click the + Add button.
4. In the pop up window, select:
   - Service:
     - Microsoft.Sql
     - Microsoft Storage (optional: only needed if you are using Cloudera Data Warehouse)
   - Subnets: Select all subnets for which you want to apply the service endpoint
5. Click Add.

Set up a service endpoint via ARM template

If you are using ARM templates to create your infrastructure, you can add the serviceEndpoints section to your template:

For example:

```
"subnets": [
        {
            "name": "<YOUR-SUBNET-NAME>",
            "properties": {
                "addressPrefix": "<YOUR-SUBNET-PREFIX>",
                "serviceEndpoints": [
                    {
                        "service": "Microsoft.Sql"
                    },
                    {
                        "service": "Microsoft.Storage"
                    }
                ]
            }
        }
    ]
```

Related Information

Private endpoint for Azure Postgres

Private endpoint for Azure Postgres

By default Cloudera uses service endpoints, but you can select to use private endpoints instead. During environment registration you can optionally select the "Create Private Endpoint" option to use private endpoints instead of using a service endpoint. Currently, only one service or private endpoint is used, for Azure Postgres.

Azure Postgres service can be reached via the following two methods, both of which are designed to allow customers to restrict who connects to the Azure Postgres service:

- Service endpoints (Created by default)

  - When a service endpoint is used, a database server must have a public IP address. This means the traffic is leaving their virtual network.
  - A service endpoint helps with the possibility of creating a firewall filter to allow connections only from the subnet explicitly linked to a given Postgres instance.
- Private endpoints (Created instead of a service endpoint only if explicitly enabled)

  - A private endpoint makes it possible to connect to an Azure Postgres server instance over a private IP address from the VNet, always ensuring traffic stays within your VNet hence increasing security.
  - A private endpoint consists of a network interface using a private IP in the VNet, and a DNS service that will resolve the FQDN of a server to the private IP address.

For more general information, see Use Virtual Network service endpoints and rules for Azure Database for PostgreSQL - Single Server and Private Link for Azure Database for PostgreSQL-Single server.

Private DNS options

Cloudera supports using private endpoints with an existing private DNS zone that can be either pre-created and provided by you, or created by Cloudera.

When using a private endpoint for Azure Postgres, an Azure private DNS zone is used for the DNS service resolving the FQDN to the private IP. Cloudera offers two options. The DNS zone can be:

- Created and managed entirely by Cloudera (you select "Create new private DNS zone" during environment registration), or
- Created by you before registering an environment (you pre-create the DNS zone and select it during environment registration). You need to provide access for discovery, validation, and adding and removing DNS A records.

Private endpoints (with either of the two DNS options) can only be used with a single customer-provided resource group. They cannot be used with Cloudera-created multiple resource groups.

You can only use private endpoints with your own private DNS zone if you specify an existing VNet for the environment. If you choose that Cloudera creates a new VNet and enables private endpoints, Cloudera is going to manage the private DNS zone for you.

Depending on whether you prefer to bring your own DNS or have Cloudera create and manage it, refer to the following documentation:

Using Cloudera-managed private DNS

Review this documentation if you are planning to use a private endpoint for Azure Postgres with a Cloudera-managed DNS.

Requirements and limitations

The following limitations apply when using a Cloudera-managed private DNS:

- Only Azure's private DNS zone is supported. Using an on-premise DNS is not supported.
- Private endpoints can only be used with a single customer-provided resource group. They cannot be used with Cloudera-created multiple resource groups. When creating an environment from the UI, the "Create Private Endpoints" option is disabled if you select the option for Cloudera to create multiple resource groups. If using CDP CLI, a validation error appears if this option is used with Cloudera creating multiple resource groups.

- The private DNS zone must be in the single existing resource group, even if the VNet is located elsewhere. This is because:

  - Microsoft requires that the DNS zone has a fixed name.
  - A VNet cannot have links to two different DNS zones with the same name.
  - It is not possible to check if a VNet already has a DNS zone with a given name attached to it.

- Only one resource group can have private endpoints with a given VNet. This is because:

  - Only one DNS zone with a given name can be linked to a VNet.
  - That DNS zone must be in the single resource group where all the resources are located.

- The private DNS zone and virtual network links are shared within the single resource group. The first environment ever created in that resource group will create them. They will never be deleted by Cloudera.
- The private DNS zone creation cannot be turned off. There is no option to create private endpoints but not create a private DNS zone. You can, however, choose to specify your own existing private DNS zone. See Bringing your own private DNS.

Prerequisites

In order to use a Cloudera-managed private DNS, you should meet the following prerequisites:

1. Disable private endpoint network policies on page 35
2. Review DNS zones existing in your resource group on page 35
3. Ensure that Cloudera has adequate permissions on page 35

Disable private endpoint network policies

Only subnets that have private endpoint network policies turned off are eligible for private endpoint creation, because network security groups (NSG) are not supported for private endpoints.

You can use the following Azure CLI command to disable private endpoint network policies for a certain subnet:

```
az network vnet subnet update \
  --name <subnet-name> \
  --resource-group <resource-group-name> \
  --vnet-name <vnet-name> \
  --disable-private-endpoint-network-policies true
```

For example:

```
az network vnet subnet update \
  --name default-2 \
  --resource-group my-cdp-rg \
  --vnet-name my-azure-vnet \
  --disable-private-endpoint-network-policies true
```

Review DNS zones existing in your resource group

If you want Cloudera to create and manage the private DNS zone, review the DNS zones that exist in the resource group that you are planning to use for Cloudera and make sure that one of the following is true:

- No private DNS zone named "privatelink.postgres.database.azure.com" is connected to the VNnet.
- If there is a private DNS zone named "privatelink.postgres.database.azure.com" connected to the VNet, verify that the zone is located in the existing resource group that you are planning to use for Cloudera. If the private DNS zone is already used for one environment, Cloudera can reuse it for another environment.

Ensure that Cloudera has adequate permissions

Ensure that the role that you are using for the Azure credential has the permissions mentioned in Role definition 2: Allows Cloudera to use only a single existing resource group and create private endpoints.

Related Information
Manage network policies for private endpoints

Bringing your own private DNS
Review this documentation if you are planning to use a private endpoint for Azure Postgres with your own private DNS.

Requirements and limitations
The following limitations apply when using your own private DNS:

- Only Azure's private DNS zone is supported. Using an on-prem DNS is not supported.
- Private endpoints can only be used with a single customer-provided resource group. They cannot be used with Cloudera-created multiple resource groups. When creating an environment from the UI, the "Create Private Endpoints" option is disabled if you select the option for Cloudera to create multiple resource groups. If using CDP CLI, a validation error appears if this option is used with Cloudera creating multiple resource groups.
- You can only use private endpoints with your own private DNS zone if you specify an existing VNet for the environment. If you chose Cloudera to create your network and enable private endpoints, then Cloudera is going to manage the private DNS zone for you.
- The private DNS zone provided to Cloudera must have the name "privatelink.postgres.database.azure.com". Furthermore, it must be in a subscription that is accessible to the service principal used by the Cloudera app-based credential; That is, the subscription where the private DNS zone is created must be in the same tenant where the service principal is located.

Prerequisites
When bringing your own private DNS, you should meet the following prerequisites:

Disable private endpoint network policies

Only subnets that have private endpoint network policies turned off are eligible for private endpoint creation, because network security groups (NSG) are not supported for private endpoints.

You can use the following Azure CLI command to disable private endpoint network policies for a certain subnet:

```
az network vnet subnet update \
   --name <subnet-name> \
   --resource-group <resource-group-name> \
   --vnet-name <vnet-name> \
   --disable-private-endpoint-network-policies true
```

For example:

```
az network vnet subnet update \
   --name default-2 \
   --resource-group my-cdp-rg \
   --vnet-name my-azure-vnet \
   --disable-private-endpoint-network-policies true
```

Create a private DNS zone and link it to your VNet

If you choose to provide your own Azure private DNS zone then you should:

1. Create a private DNS zone with the name "privatelink.postgres.database.azure.com" in any subscription accessible to the service principal used by the Cloudera app-based credential; That is, the subscription where the private DNS zone is created must be in the same tenant where the service principal is located.

   To create a private DNS zone, you can use the following Azure CLI command:

   ```
   az network private-dns zone create \
        --name privatelink.postgres.database.azure.com \
        --resource-group <YOUR_DNS_RESOURCE_GROUP> \
   ```

```
        --subscription <YOUR_SUBSCRIPTION_FOR_DNS>
```

2. Link it to the VNet that you are planning to use for the Cloudera environment.

You can do this using the following Azure CLI command:

```
az network private-dns link vnet create \
    --name <DESIRED_LINK_NAME> \
    --resource-group <YOUR_VNET_RESOURCE_GROUP> \
    --zone-name privatelink.postgres.database.azure.com \
    --virtual-network <YOUR_VNET_RESOURCE_ID> \
    --subscription <YOUR_VNET_SUBSCRIPTION>
```

Ensure that Cloudera has adequate permissions

Ensure that the role that you are using for the Azure credential has the permissions mentioned in Role definition 2: Allows Cloudera to access and use only a single existing resource group and create private endpoints.

Additionally, if your DNS zone is outside of the single resource group that you are planning to provide to Cloudera, the following additional permissions are required for the service principal to be able to discover, validate, and use the DNS zone (add and remove A records). These additional permissions need to be included in a separate role definition and assigned separately.

The placeholders in bold must be replaced with actual values:

- {SUBSCRIPTION-ID} - The ID of the subscription where the DNS zone is located.
- {RESOURCE-GROUP-NAME} - The name of the resource group where the DNS zone is located.

```
{
    "Name": "Cloudera Management Console Azure Operator for Using Private
DNS Zones",
    "IsCustom": true,
    "Description": "Can list, validate and use private DNS zones",
    "Actions": [
      "Microsoft.Network/privateEndpoints/privateDnsZoneGroups/read",
      "Microsoft.Network/privateEndpoints/privateDnsZoneGroups/write",
      "Microsoft.Network/privateDnsZones/join/action",
      "Microsoft.Network/privateDnsZones/read",
      "Microsoft.Network/privateDnsZones/virtualNetworkLinks/read"
    ],
    "NotActions": [],
    "DataActions": [
    ],
    "NotDataActions": [],
    "AssignableScopes": [
      "/subscriptions/{SUBSCRIPTION-ID}/resourcegroups/{RESOURCE-GROUP-
NAME}"
    ]
  }
```

Related Information
Manage network policies for private endpoints
Quickstart: Create an Azure private DNS zone using the Azure portal

Resources created under the hood
Creating a Cloudera environment on Azure with private endpoints in Cloudera involves creating several Azure resources, all of which are necessary to have a private endpoint setup that works out-of-the-box.

These create a setup where:

- A private endpoint to the Postgres server can be reached.
- A network interface with only a private IP address exists in the subnet where the private endpoint has been added.

This ensures that communication between Cloudera and the Azure Postgres server happens via a private IP address. However, cluster services do need to contact the Postgres server via FQDN, so the address needs to be resolvable from the VNet.

Resources created when using a Cloudera-managed private DNS zone

In the scenario where Cloudera creates the private DNS, several resources are created:

- A private DNS zone: It is a DNS zone, part of an Azure-hosted DNS server. It has a fixed name ("privatelink.postgres.database.azure.com").
- A virtual network link between the zone and a VNet where the domain resolution should happen.
- An "A" record within the zone: FQDN to IP address resolution. There is no reverse lookup.
- Since communication can flow over a private IP address, Cloudera sets "Deny public access" on the Azure Postgres server.

> **Note:**
>
> If there is a private DNS zone named "privatelink.postgres.database.azure.com" containing the required virtual network link in the deployment's resource group, Cloudera reuses it. The "A" record and the network link are created only if they do not already exist.

When you delete the environment, the DNS zone and the network link that you provided will not be deleted.

Resources created when using your own private DNS zone

In this case the private DNS zone and the network link are provided by you and Cloudera creates the following:

- Cloudera creates an "A" record within the zone: FQDN to IP address resolution. There is no reverse lookup.
- Since communication can flow over a private IP address, Cloudera sets "Deny public access" on the Azure Postgres server.

When you delete the environment, the DNS zone and the network link that you provided will not be deleted.

Related Information

Virtual network workloads without custom DNS server

Azure services DNS zone configuration

Network security groups

Network security groups (NSGs) determine the inbound and outbound traffic to and from your Cloudera environment. That is, you should use security group settings to allow users from your organization access to Cloudera resources.

You have two options:

- Use your existing security groups (recommended for production)
- Have Cloudera create new security groups

You should verify the security group limits in your Azure account to ensure that you can create security groups for Cloudera.

Existing security groups

If you would like to create your own security groups, two security groups need to be created: the first security group will be used for all gateway nodes and the second security groups will be used for all other nodes. The gateway nodes communicate with the Cloudera Management Console and therefore require additional ports. These security groups will be applied when creating a data lake and FreeIPA during environment creation and when you create Cloudera Data Hub clusters.

Review the following guidelines prior to adding security groups rules. This describes all the inbound ports that need to be open and provides guidelines for what to enter as a source range:

Note: The communication via TCP/UDP 0-65535 and ICMP is essential for healthy operation of CDP environments, Data Hubs, and data services running within the VNet, so ensure that you open these ports as described below. While some services only need well-known fixed ports, a majority of them depend on ephemeral (i.e. dynamically or randomly allocated) ports; This is why the wildcard 0-65535 TCP/UDP port range is used in the absence of a detailed breakdown of individual ports. Since overall access to the VNet is typically secured by other means, the use of the wildcard rules does not pose a higher risk against external attacks.

"Knox" security group

This is used for gateway nodes:

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts. You should open it to your organization's CIDR.
TCP	443	Your CIDR and CDP CIDR	This port is used to access the Data Lake and Data Hub cluster UIs via Knox gateway. You must open this port to your organization's CIDR in order to access cluster UIs. When CCM is enabled, you only need to set this to your CIDR.
TCP	9443	CDP CIDR	This port is used by CDP to maintain management control of clusters and data lakes. By default, when CDP creates the security groups automatically, it opens this port to the correct IP. This port is not needed when CCM is enabled.
TCP, UDP	0-65535	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

"Default" security group

This is used for all nodes except Knox gateway nodes:

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to the hosts. You should open it to your organization's CIDR.
TCP	443	Your CIDR	This port is only required if you are planning to spin up Machine Learning workspaces since HTTPS access to ML workspaces is available over port 443. If you are not planning to use the Machine Learning service, you do not need to open this port.

Protocol	Port Range	Source	Description
TCP	9443	CDP CIDR	This port is used by CDP to maintain management control of clusters and data lakes. By default, when CDP creates the security groups automatically, it opens this port to the correct IP. This port is not needed when CCM is enabled.
TCP, UDP	0-65535	Your VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet. TCP port 5432 is used by the Data Lake for communication with its attached database.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Security groups can be created and managed from the Azure Portal > Network Security Groups. For detailed instructions on how to create new security groups on Azure, refer to Filter network traffic with a network security group using the Azure portal.

On the Network Interface page > Settings, click Network security group.

On the Network Security Groups page, click Inbound Security Rules to view the list of rules.

In the Inbound security rules tab, click Add.

Add inbound security rule
masterdp-az-dl12169sg

Basic

You need to create two security groups: Knox and Default (You will see this terminology in the Cloudera Management Console UI and CLI, so if you decide to choose different names, make sure that you are able to distinguish between the two security groups).

Note:

There is a known issue where even if you create and specify your own security groups the Cloudera Data Warehouse and Cloudera AI services create their own security groups. For instructions on how to restrict access on the security groups created by the Cloudera Data Warehouse service, refer to Restricting access to endpoints in AWS environments.

New security groups

If you would like Cloudera to create the security groups for you, you need to provide a CIDR range for inbound traffic to Azure instances from your organization. Cloudera creates multiple security groups: one for each Data Lake host group, one for each FreeIPA host group, and one per host group when Cloudera Data Hub, Cloudera Data Warehouse, and Cloudera AI clusters are created. On these security groups, Cloudera opens ports as described in Default security group settings documentation.

Note:

There is a known issue where even if you create and specify your own security groups, the Cloudera Data Warehouse and Cloudera AI services create their own security groups. For instructions on how to restrict access on the security groups created by the Cloudera Data Warehouse service, refer to Restricting access to endpoints in AWS environments.

Related Information

Restricting access for Cloudera services that create their own security groups on Azure

Default security group settings on Azure

Depending on what you chose during environment creation, Cloudera can create security groups for your environment automatically or you can provide your own security groups.

Note:

Even if you create and specify your own security groups, the Cloudera Data Warehouse and Cloudera AI services create their own security groups. Refer to Restricting access for Cloudera services that create their own security groups on Azure for instructions on how to restrict access.

Note: The communication via TCP/UDP 0-65535 and ICMP is essential for healthy operation of CDP environments, Data Hubs, and data services running within the VNet, so ensure that you open these ports as described below. While some services only need well-known fixed ports, a majority of them depend on ephemeral (i.e. dynamically or randomly allocated) ports; This is why the wildcard 0-65535 TCP/UDP port range is used in the absence of a detailed breakdown of individual ports. Since overall access to the VNet is typically secured by other means, the use of the wildcard rules does not pose a higher risk against external attacks.

Environment security groups

Depending on what you chose during environment creation, Cloudera can create security groups for your environment automatically or you can provide your own security groups.

- If you choose to use your own security groups, you are asked to create Knox and Default security groups as described in the Security groups documentation.
- If you choose for Cloudera to create all security groups required for an environment, the following security groups are created:

Data Lake: master

Azure naming convention: master${dl-name}${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts. You should open it to your organization's CIDR.
TCP	443	Your CIDR and Cloudera CIDR	This port is used to access the Data Lake and Cloudera Data Hub cluster UIs via Knox gateway. You should open it to your organization's CIDR in order to access cluster UIs. This port is also required if you are planning to spin up Cloudera AI workbenches since HTTPS access to Cloudera AI workbenches is available over port 443. If you are not planning to use the Cloudera AI service, you do not need to open this port. When Cluster Connectivity Manager is enabled, you only need to set this to your CIDR.
TCP	9443	Cloudera CIDR	This port is used by Cloudera to maintain management control of clusters and data lakes. This port is not used when Cluster Connectivity Manager is enabled.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Data Lake: IDBroker

AWS naming convention: ${environment-name}-${random-id}-ClusterNodeSecurityGroupidbroker-${random-id}

Azure naming convention: idbroker${dl-name}${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

FreeIPA

Azure naming convention: master0-${env-name}freeipa${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts. You should open it to your organization's CIDR.

Protocol	Port Range	Source	Description
TCP	9443	Cloudera CIDR	This port is used by Cloudera to maintain management control of clusters and data lakes. This port is not used when Cluster Connectivity Manager is enabled.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Cloudera Data Hub security groups

Depending on what you chose during environment creation, Cloudera can create security groups for your Data Hub clusters automatically or it can use your pre-created security groups:

- If during environment creation, you provided your own security groups, Cloudera uses these security groups when deploying clusters.
- If during environment creation you chose for Cloudera to create new security groups, new security groups are created for each Cloudera Data Hub cluster as follows:

Cloudera Data Hub: master

Azure naming convention: ${hostgroup-name}${dh-name}${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts. You should open it to your organization's CIDR.
TCP	443	Your CIDR and Cloudera CIDR	This port is used to access the Data Lake and Cloudera Data Hub cluster UIs via Knox gateway. You should open it to your organization's CIDR in order to access cluster UIs. When Cluster Connectivity Manager is enabled, you only need to set this to your CIDR.
TCP	9443	Cloudera CIDR	This port is used by Cloudera to maintain management control of clusters and data lakes. This port is not used when Cluster Connectivity Manager is enabled.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Cloudera Data Hub: worker

Azure naming convention: ${hostgroup-name}${dh-name}${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Cloudera Data Hub: compute

Azure naming convention: ${hostgroup-name}${dh-name}${numeric-id}sg

Protocol	Port Range	Source	Description
TCP	22	Your CIDR	This is an optional port for end user SSH access to cluster hosts.
TCP, UDP	0-65535	Your VNet's CIDR (for example 10.10.0.0/16) and your subnet's CIDR (for example 10.0.2.0/24).	This is required for internal communication within the VNet.
ICMP	N/A	Your internal VNet CIDR (for example 10.10.0.0/16).	This is required for internal communication within the VNet.

Cloudera Data Warehouse security groups

Cloudera always creates new security groups when Cloudera Data Warehouse are deployed.

Cloudera AI security groups

Cloudera always creates new security groups when Cloudera AI workbenches are deployed.

Cloudera Data Engineering security groups

Cloudera always creates new security groups when Cloudera Data Engineering clusters are deployed.

Cloudera DataFlow security groups

Cloudera always creates new security groups when Cloudera DataFlow environments are enabled.

SSH key pair

When registering an environment, you will be asked to upload a BC FIPS SSH key pair. The minimum SSH key size is 4096 bits.

The SSH key will be used for root-level access to Data Lake and Cloudera Data Hub instances.

If you need to generate an SSH key, you can use typical SSH key generation steps, such as those described in Quick steps: Create and use an SSH public-private key pair for Linux VMs in Azure .

Virtual machines

Cloudera provisions Virtual machines (VMs) as part of the environment creation process (for Data Lake and FreeIPA) and for compute clusters.

Therefore, you should verify the limits on the number and type of VM instances in your Azure account to ensure that you are able to provision an environment and create clusters in Cloudera.

Cloudera supports Azure reserved VM instances; That is, If you have purchased reserved instances, Cloudera uses them automatically according to Azure policy.

Cloudera provides default Virtual Hard Drives (VHDs) and provisions managed images based on these VHDs. These managed images are used for Data Lake, FreeIPA, and compute cluster instances.

For a list of supported VM types, refer to Cloudera on cloud service rates.

Custom images

By default Cloudera provides a set of default images that are used for all provisioned VMs, but you can optionally use custom images for Data Lake, FreeIPA, and Cloudera Data Hub.

You might require a custom image for compliance or security reasons (a "hardened" image), or to have your own packages pre-installed on the image, for example monitoring tools or software.

If you would like to use custom images instead of the default images, refer to Custom images and image catalogs.

ADLS Gen2 and managed identities for logs, backups, and data storage

Cloudera requires that you create and provide an Azure Data Lake Store Gen 2 (ADLS Gen2) storage account and create a container within it for storing workload data and logs. Furthermore, in order for Cloudera to be able to access this container, you must create and assign managed identities.

You must provide the following:

- Create an ADLS Gen2 storage account with hierarchical namespace enabled. The storage account must be in the same region as the environment.
- You must create and provide multiple user-assigned managed identities that allow access to the ADLS Gen2 location.

Cloudera supports both standard and premium ADLS Gen2 storage accounts.

For more information about the ADLS Gen2 container and managed identities setup, refer to the following documentation:

Minimal setup for Azure cloud storage

This minimal secure setup uses one ADLS Gen2 storage account with multiple containers in it, and multiple managed identities where each managed identity has at least one role assigned.

ADLS Gen2 storage account

You should create one ADLS Gen2 storage account with two containers within it (one for Storage Location Base and another for Logs Location Base). Additionally, you can specify a container for Backup Location Base to store FreeIPA and Data Lake backup data separately from logs:

- One ADLS Gen2 container is required to use as Storage Location Base such as abfs://storagefs@mydatalake. dfs.core.windows.net where mydatalake is your storage account name and storagefs is your container name. The Storage Location Base is used for storing workload data and Ranger audits.
- One ADLS Gen2 container is required to use as Logs Location Base such as abfs://logsfs@mydatalake.dfs.core.wi ndows.net where mydatalake is your storage account name and logsfs is your container name. The Logs Location Base is used for Data Lake, FreeIPA and Data Hub logs, and FreeIPA and Data Lake backups.

  Note: Ranger audits are stored under Storage Location Base and not under Logs Location Base. The Logs Location Base is used for Data Lake, Data Hub and FreeIPA logs and, if no separate container is provided, FreeIPA backups.

- (Optional) One optional ADLS Gen2 container to use as Backup Location Base such as abfs://backupfs@myda talake.dfs.core.windows.net where mydatalake is your storage account name and backupfs is your container name. The Backup Location Base is used for FreeIPA and Data Lake backups. If a separate container is not provided, the backups are stored in the Logs Location Base.

Storage Location Base examples

	abfs://storagefs@mydatalake.dfs.core.windows.net
Ranger Audit Logs	abfs://storagefs@mydatalake.dfs.core.windows.net/ranger/audit

Logs Location Base examples

	abfs://logsfs@mydatalake.dfs.core.windows.net
FreeIPA Logs	abfs://logsfs@mydatalake.dfs.core.windows.net/cluster-logs/freeipa If your environment was created prior to February 2021, this is abfs://logsfs@mydatalake.dfs.core.windows.net/freeipa

Backup Location Base examples

If you specify a separate container for FreeIPA backups, the backups are written to that container:

	abfs://backupfs@mydatalake.dfs.core.windows.net
FreeIPA Backup	abfs://backupfs@mydatalake.dfs.core.windows.net/cluster-backups/freeipa

If the separate container is not provided, the FreeIPA backups are written to the Logs Location Base. In both cases, the same cluster-backups/freeipa directory structure is created within the container.

Managed identities

You should create four managed identities.

The IDBroker component of Cloudera uses user-assigned managed identities for controlling access to ADLS Gen2 and stores and manages the mappings between the services/users and the corresponding managed identities. The following managed identities must be created:

Managed identity	Description	Steps
Assumer (Required)	During Data Lake cluster creation, Cloudera attaches this identity to the IDBroker VM. IDBroker then uses it to attach the other managed identities to the IDBroker VM. Once these identities are attached to the VM, IDBroker can acquire an access token for them (to eliminate the need to store credentials in the application).	1. Create a managed identity. 2. Assign the Virtual Machine Contributor and Managed Identity Operator roles to this managed identity on the scope of either the subscription or resource group. 3. Assign the Storage Blob Data Contributor role to this managed identity on the scope of the Logs Location Base and Backup Location Base (if created) containers created for Cloudera.
Data Lake Admin (Required)	This managed identity is used for Cloudera services to access data.	1. Create a managed identity. 2. Assign the Storage Blob Data Owner role to this managed identity on the scope of the three containers (Storage Location Base, Logs Location Base, and Backup Location Base if it exists) created for Cloudera.

Managed identity	Description	Steps
Ranger Audit Logger (Required)	This managed identity is used by Ranger to write audits.	1. Create a managed identity. 2. Assign the Storage Blob Data Contributor role to this managed identity on the scope of the Storage Location Base container created for Cloudera. 3. Additionally, if you created a separate container for Backup Location Base, assign the Storage Blob Data Contributor role to this managed identity on the scope of the Backup Location Base. Otherwise, assign it on the Logs Location Base container.
Logger (Required)	This managed identity is used by Cloudera to write telemetry logs.	1. Create a managed identity. 2. Assign Storage Blob Data Contributor role to this managed identity on the scope of the Logs Location Base and Backup Location Base (if created) created for Cloudera when using lower versions than Cloudera Public Cloud 7.2.17. ⚠ **Important:** If you are using Cloudera AI, you will also need the Storage Account Contributor role for the data lake storage account assigned to the Logger managed identity. For more information about permissions for Cloudera AI on Azure, see Cloudera AI minimum permissions.
Ranger RAZ (Optional)	This managed identity is only required if you are planning to use Fine-grained access control. It is used by Cloudera to orchestrate fine-grained access control to your data storage location.	1. Create a managed identity. 2. Assign the Storage Blob Data Owner role and the Storage Blob Delegator to this managed identity on the scope of the ADLS Gen2 storage account created for Cloudera. If you prefer not to assign the Storage Blob Data Owner role, you can Create a custom role to use in RAZ-enabled Azure environment and assign that role and the Storage Blob Delegator instead.

The following diagram illustrates the minimal required setup where Backup Location Base is in the same location as the Logs Location Base:

The following diagram illustrates the minimal required setup where Backup Location Base and Logs Location Base are separate:

If you are planning to use fine-grained access control, one additional managed identity is required, as illustrated in the following diagrams.

The first diagram illustrates the scenario where Backup Location Base is in the same location as the Logs Location Base:

The second diagram illustrates the scenario where Backup Location Base and Logs Location Base are separate:

The following documentation provides detailed steps on how to create this setup. The steps involve:

1. Creating an ADLS Gen2 storage account and containers
2. Creating managed identities for the minimal setup

3. Providing the parameters in the Cloudera UI

Creating ADLS Gen2 storage account and containers

Create a resource group and then create a storage account and two containers within it. The storage account must have the hierarchical namespace enabled.

1. First, create a resource group that can act as a logical grouping of storage accounts. To create the resource group, follow the steps described in Create resource groups in Azure docs. Make sure to create it in the specific region that you would like to use for your environment.

2. Next, create an ADLS Gen2 storage account. In our example setup, the storage account name is my-datalake. To create the storage account, follow the Create an account using the Azure portal in Azure docs. In the Advanced > Data Lake Storage Gen2 section, make sure to Enable hierarchical namespace:

> **Note:** You can reuse the same resource group that you created for the storage account or you can optionally create a new resource group that can act as a logical grouping of managed identities.

3. After creating an ADLS Gen 2 storage account, create two containers within it (one for Storage Location Base and another for Logs Location Base). You can also optionally create a third container for Backup Location Base.

To create a container, follow the usual steps:

   a. On Azure Portal, navigate to Storage Accounts > your newly created storage account > Containers > +Container.

   b. Provide a name for your container and click OK.

Repeat these steps to create all required containers. In our example setup, the containers are called storagefs, logs fs, and backupfs.

Once you have created the storage account and container, note the created resources in the following format:

abfs://$[container-name]@$[storage-account-name].dfs.core.windows.net

For example, in our example setup, we created the following two containers:

abfs://storagefs@mydatalake.dfs.core.windows.net

abfs://logsfs@mydatalake.dfs.core.windows.net

abfs://backupfs@mydatalake.dfs.core.windows.net

Creating Azure managed identities

Once you've created the storage account and file system within it, create the managed identities and then assign roles with specific scopes to these identities.

You can reuse the same resource group that you created for the storage account or you can optionally create a new resource group that can act as a logical grouping of managed identities.

You need to create four or five managed identities (Assumer, Data Lake Admin, Ranger Audit Logger, Logger, and optionally the Ranger RAZ). Use the following steps to create these managed identities:

1. On Azure Portal, navigate to Managed Identities.
2. Click +Add.
3. Specify managed identity name and select the resource group that you created earlier.

Repeat these steps to create each of the four managed identities. Once you've created these managed identities, assign roles with specific scopes (subscription or storage account) to these identities as follows.

Create the Assumer identity

Assign the Virtual Machine Contributor role and the Managed Identity Operator role to the Assumer managed identity on either the subscription or resource group level, and then assign the Storage Blob Data Contributor on the scope of the container created earlier for Logs Location Base.

Steps

1. Navigate to Subscriptions > your subscription > Access Control (IAM).
2. Click +Add.
3. Under Add role assignment:

   a. Under Role, select Virtual Machine Contributor.
   b. Under Assign access to, select User assigned managed identity.
   c. Under Select, select the Assumer managed identity created earlier.
   d. Click Save.
4. Repeat the role assignment steps 2-3, but this time assign the Managed Identity Operator role to the Assumer Identity.

Next, assign the Storage Blob Data Contributor role to the Assumer managed identity on the scope of the containers created earlier for Logs Location Base and Backup Location Base (if separate):

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).
2. Click +Add > Add role assignment.
3. Under Add role assignment:

   a. Under Role, select Storage Blob Data Contributor.
   b. Under Assign access to, select User assigned managed identity.
   c. Under Select, select the Assumer managed identity created earlier.
   d. Click Save.

Create the Data Lake Admin identity

Assign the Storage Blob Data Owner role to the Data Lake Admin managed identity on the scope of the two containers created earlier for Storage Location Base and Logs Location Base, and to the Backup Location Base container if it exists. You need to do this separately for each of the containers.

Steps

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).
2. Click +Add > Add role assignment.

3. Under Add role assignment:

    a. Under Role, select Storage Blob Data Owner.

    b. Under Assign access to, select User assigned managed identity.

    c. Under Select, select the Data Lake Admin managed identity created earlier.

    d. Click Save.

Repeat the steps for the second container and for the third container (if you created it).

Create the Ranger Audit Logger identity

Assign the Storage Blob Data Contributor role to the Ranger Audit Logger managed identity on the scope of the container created earlier for Storage Location Base. If you created a separate container for Backup Location Base, you should also repeat these steps on the container created earlier for Backup Location Base. Otherwise, they should be repeated on the Logs Location Base container.

Steps

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).

2. Click +Add > Add role assignment.

3. Under Add role assignment:

    a. Under Role, select Storage Blob Data Contributor.

    b. Under Assign access to, select User assigned managed identity.

    c. Under Select, select the Ranger Audit Logger managed identity created earlier.

    d. Click Save.

Repeat the steps for the second container and for the third container (if you created it).

Create the Logger identity

Assign the Storage Blob Data Contributor role to the Logger managed identity on the scope of the container created earlier for Logs Location Base. If you created a separate container for Backup Location Base, you should also assign the same role to the same managed identity on the scope of the container created earlier for Backup Location Base.

Steps

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).

2. Click +Add > Add role assignment.

3. Under Add role assignment:

    a. Under Role, select Storage Blob Data Contributor.

    b. Under Assign access to, select User assigned managed identity.

    c. Under Select, select the Logger managed identity created earlier.

    d. Click Save.

4. If you created a separate container for Backup Location Base, you should perform steps 1-3 to also assign the Storage Blob Data Contributor role to the Logger managed identity on the scope of the container created earlier for Backup Location Base.

Create the Ranger RAZ identity (Optional)

If you would like to use Fine-grained access control, you should create the Ranger RAZ managed identity. Assign the Storage Blob Data Owner role and the Storage Blob Delegator role to the Ranger RAZ managed identity on the scope of the storage account that you created earlier (which contains the Storage Location Base and Logs Location Base).

Steps

1. Navigate to Storage accounts > your storage account > Access Control (IAM).

2. Click +Add > Add role assignment.

3. Under Add role assignment:

- **a.** Under Role, select Storage Blob Data Owner.
- **b.** Under Assign access to, select User assigned managed identity.
- **c.** Under Select, select the Ranger RAZ managed identity created earlier.
- **d.** Click Save.

4. Repeat the role assignment steps 2-3, but this time assign the Storage Blob Delegator role to the Ranger RAZ managed identity.

After performing these steps, you should have the required managed identities created and their roles assigned on the correct scope.

Providing the parameters in the Cloudera UI

Once you've created the ADLS Gen2 location and the required managed identities, provide the information related to these resources in the Register Environment wizard.

Parameter	Description
Data Access and Audit	
Assumer Identity	Select the Assumer identity created earlier.
Storage Location Base	Enter the Storage Location Base created earlier. In our example this was abfs://storagefs@mydatalake.dfs.core.windows.net, so you should enter storagefs@mydatalake.
Data Access Identity	Select the Data Lake Admin identity created earlier.
Ranger Audit Identity	Select the Ranger Audit identity created earlier.
Fine-grained access control on ADLS Gen2	
Enable Ranger authorization for ADLS Gen2	If you would like to use Fine-grained access control, click on "Enable Ranger authorization for ADLS Gen2". Next, select the Ranger RAZ identity created earlier.
Select Azure managed identity for Ranger authorization	
Logs	
Logger Identity	Select the Logger identity created earlier.
Logs Location Base	Enter the Logs Location Base created earlier. In our example this was abfs://logsfs@mydatalake.dfs.core.windows.net, so you should enter logsfs@mydatalake.
Backup Location Base (Optional)	If you created it, enter the Backup Location Base. In our example this was abfs://backupfs@mydatalake.dfs.core.windows.net, so you should enter backupfs@mydatalake. This is optional. If you don't provide this, FreeIPA and Data Lake backups will be stored in the Logs Location Base.

Onboarding Cloudera users and groups for Azure cloud storage (RAZ environments)

If your Azure environment has Fine-grained access control enabled, you should onboard your users using Ranger instead of using IDBroker.

For more information, refer to Using Ranger to Provide Authorization in ClouderaUsing Ranger to Provide Authorization in Cloudera.

Onboarding Cloudera users and groups for Azure cloud storage (No RAZ)

The minimal setup defined earlier spins up a Cloudera environment and Data Lake with no end user access to cloud storage. Adding users and groups to a Cloudera environment involves ensuring that they are properly mapped to managed identities to access cloud storage.

Note: If you are using Fine-grained access control, you should onboard your users using Ranger instead of using IDBroker mappings. Adding IDBroker mappings is disabled for RAZ-enabled environments.

In general, to have new users or groups onboarded, you need to have the following pre-created in Azure:

1. First, you need to create two more containers within the storage account (mydatalake) created earlier, one for data engineers (for example, dataeng) and one for data scientists (for example, datascience).
2. Next, you need to create two more managed identities, one for data engineers (for example, data-eng-mi) and one for data scientists (for example, data-science-mi) and assign the Storage Blob Data Owner role on the scope of one these two nearly created containers. The data-eng-mi managed identity will need the Storage Blob Data Owner role on the scope of the dataeng container and the data-science-mi managed identity will need the Storage Blob Data Owner role on the scope of the datascience container.
3. Finally, you also need to grant the Data Lake Admin managed identity created earlier the Storage Blob Data Owner role on the scope of these two newly created containers.

The final goal is to have the following that builds on the minimal setup presented earlier.

The first diagram illustrates the scenario where Backup Location Base is in the same location as the Logs Location Base:

The second diagram illustrates the scenario where Backup Location Base and Logs Location Base are separate:

The following documentation provides detailed steps for how to create this setup. The steps involve:

1. Creating additional containers
2. Crating additional managed identities
3. Creating mappings in Cloudera
4. Updating the Data Lake Admin managed identity

Creating additional containers

Within the ADLS Gen 2 storage account (mydatalake) created earlier, create two more containers (dataeng and data science).

To create a file system, perform the following steps:

1. On Azure Portal, navigate to Storage Accounts > your newly created storage account > Containers > +Container.
2. Provide a name for your container and click OK.

Repeat these steps to create both containers.

Creating additional managed identities

After creating the two additional container, create two additional managed identities, one for data engineers (data-eng -mi) and one data scientists (data-science-mi).

To create the two new managed identities, perform the following steps:

1. On Azure Portal, navigate to Managed Identities.
2. Click +Add.
3. Specify managed identity name and select the resource group that you created earlier.

Repeat these steps to create each of the two managed identities. Once you've created these managed identities, assign roles with specific scopes (limited to one of the two containers, dataeng or datascience respectively) to these identities as follows:

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).

2. Click +Add > Add role assignment.
3. Under Add role assignment:

    a. Under Role, select Storage Blob Data Owner.

    b. Under Assign access to, select User assigned managed identity.

    c. Under Select, select the managed identity.

    d. Click Save.

After performing these steps for each of the two managed identities, you should have the required managed identities created and their roles assigned on the correct scope.

Adding Cloudera user/group to managed identity mappings

After creating the two additional managed identities, one for data engineers (data-eng-mi) and one data scientists (data-science-mi), map them to specific user/group in Cloudera.

> **Note:** If you are using Fine-grained access control, you should onboard your users using Ranger instead of using IDBroker mappings. Adding IDBroker mappings is disabled for RAZ-enabled environments.

> **Note:**
>
> If a user is mapped to multiple roles via group membership, the specific role to be used needs to be provided at runtime. If the user is mapped directly to a role, the direct mapping takes precedence over mapping via group membership. For information on how to specify the role, refer to Specifying a group when user belongs to multiple groups.

Required role: DataSteward, EnvironmentAdmin, or Owner

Steps

For Cloudera UI

1. The option to add/modify these mappings is available from the Management Console under Environments > click on an environment > Actions > Manage Access > IDBroker Mappings > Edit.
2. Under Current Mappings, click Edit.
3. Click + to display a new field for adding a mapping.
4. Provide the following:

    a. The User or Group dropdown is pre-populated with Cloudera users and groups. Select the user or group that you would like to map.

    b. Under Role, specify the resource ID of a managed identity (copied from Azure Portal). You should select your data-eng-mi here.

5. Repeat the previous two steps to add additional mapping for the data-science-mi.
6. Click Save and Sync.

For CDP CLI

If you would like to create the mappings via CDP CLI, you can:

1. Use the cdp environments get-id-broker-mappings command to obtain your current mappings.
2. Use the cdp environments set-id-broker-mappings command to set additional mappings. The only way to use this command is to:

- Pass all the current mappings
- Add the new mappings

3. Next, sync IDBroker mappings. For example:

```
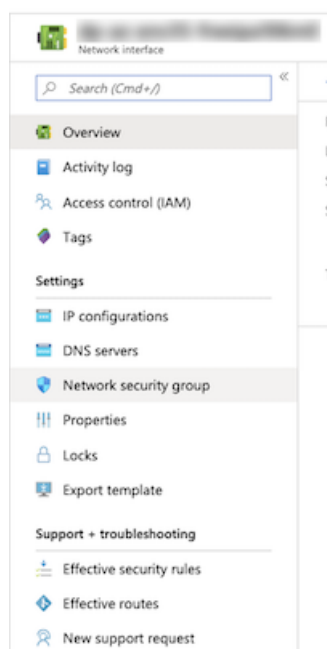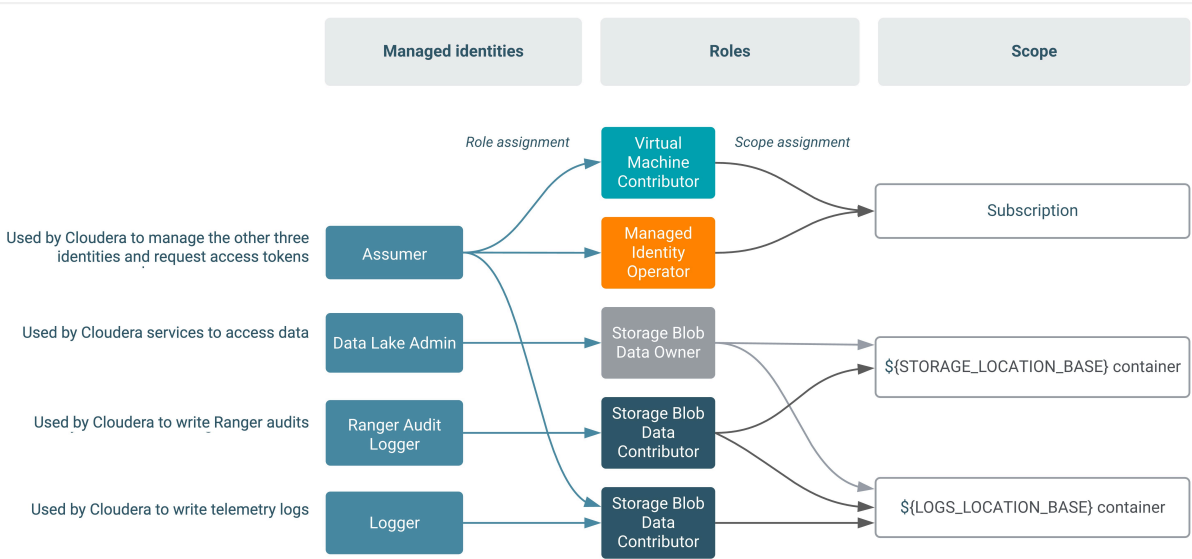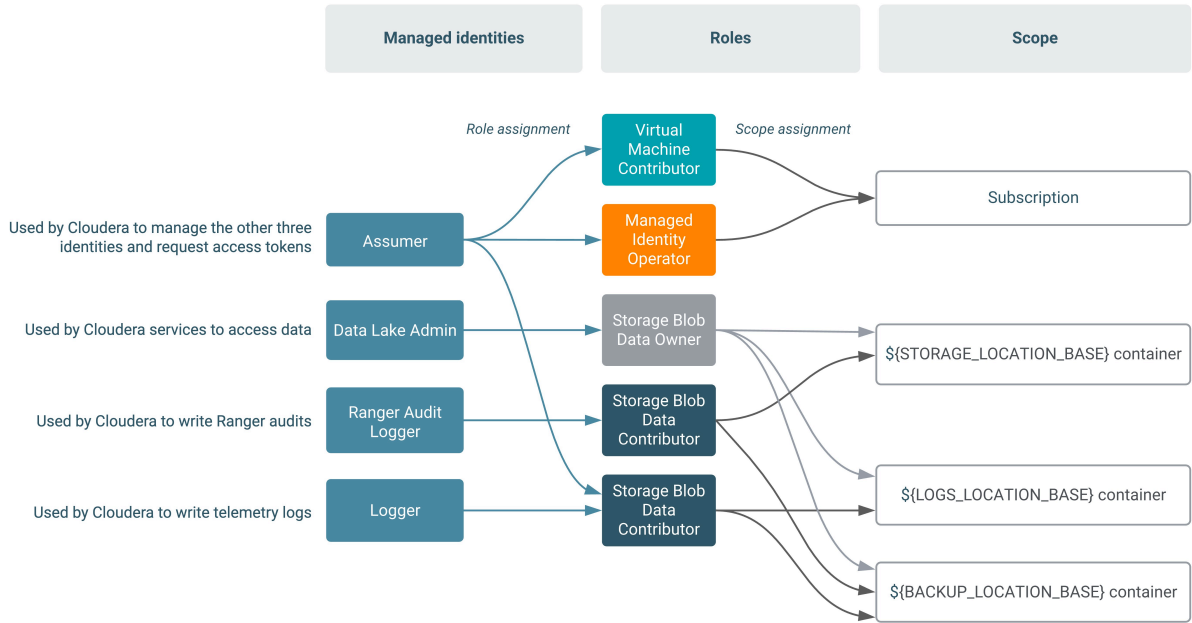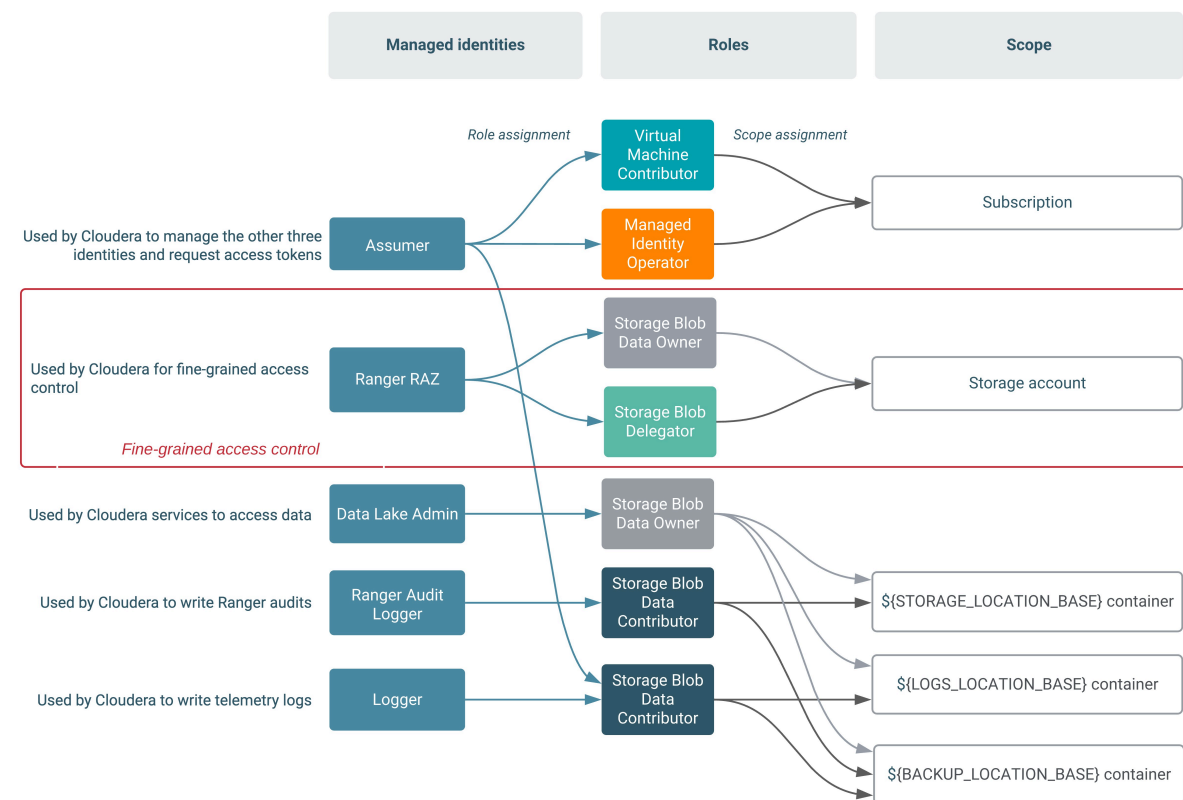cdp environments sync-id-broker-mappings --environment-name demo3
```

4. Finally, check the sync status. For example:

```
cdp environments get-id-broker-mappings-sync-status --environment-name d
emo3
```

Updating access for the Data Lake Admin managed identity

Grant the Data Lake Admin identity created earlier the Storage Blob Data Owner role on the scope of the two newly created containers.

Perform the following steps for both dataeng or datascience containers to grant the Data Lake Admin managed identity access to them:

1. Navigate to Storage accounts > your storage account > Containers > your container > Access Control (IAM).
2. Click +Add > Add role assignment.
3. Under Add role assignment:

    a. Under Role, select Storage Blob Data Owner.
    b. Under Assign access to, select User assigned managed identity.
    c. Under Select, select the Data Lake Admin Identity created earlier.
    d. Click Save.

Repeat these steps to provide Data Lake admin with access to both containers.

Using ADLS Gen2 encryption

By default ADLS Gen2 uses TLS. If you use the fs.azure.always.use.https property to turn off this behavior, you must specify abfss as the prefix in the URI to use TLS. Otherwise, you can use abfs.

ADLS Gen2 account for storing operating system images

Cloudera uses an ADLS Gen2 storage account for storing operating system images used for VMs. By default, Cloudera creates this account during environment registration, but you can optionally pre-create it. If needed, you can also copy the VHD files and create image resources manually.

Cloudera uses an ADLS Gen2 for storing operating system images used for VMs. These images are used for:

• Data Lake and Cloudera Data Hub clusters (A single image is used for both)
• FreeIPA (A separate image is used)

For each of these, there is one image for each Runtime version and for each Azure region.

Preparing images for VMs on Azure is a three-step process, involving the following steps:

1. Creating a new storage account in your subscription. The storage account is reachable from all networks so that Cloudera can access it (although the storage container is private).
2. Copying one or more virtual hard disks (VHDs) from Cloudera's regional storage account to your storage account created earlier.
3. Creating an image resource from each copied VHD.

By default, these steps are performed automatically by Cloudera during the "Setting up Cloudera image" and the "Creating infrastructure" steps of launching a stack. However, if due to your organization's security policies, you do not want Cloudera to create your storage account or you do not want the storage account to be publicly available, you have two alternatives:

Scenario	Solution	Steps
You do not want Cloudera to create a storage account in your Azure subscription or Cloudera is not granted the permission to create new storage accounts in your Azure subscription, but you are willing to provide Cloudera access to the storage account once it exists.	Manually pre-create a storage account that is available from all networks, and once it exists, Cloudera can copy VHD files and create the image resources.	Refer to Creating the storage account for images manually.
Image copying by Cloudera is not possible because the storage account should not be publicly available.	You should perform all three steps manually.	See all three steps linked in this section.

> **Note:**
>
> The storage account and the VHDs themselves are not deleted upon deleting the environment to save time when creating future environments.

If you would like to perform these manual steps, refer to the following instructions.

Creating a storage account for operating system images manually

In some cases where your organization's security policies require it, you may want to manually pre-create the storage account that Cloudera uses for storing operating system images.

Steps

1. Create a resource group on Azure with the name "cloudbreak-images".
2. Create a storage account on Azure.

   a. The name is the concatenation of the following elements. It is different depending on whether or not you would like to use your existing resource group for Cloudera:

   ### Table 1:

	Naming convention
Single existing resource group	Enter a name that is a simple concatenation of the following four elements: • "cbimg" • Region identifier: The starting letters of the region where the SA is. For example, "eu" for East US, or "eu2" for East US 2 • The adler32 checksum without leading zeroes of the Subscription ID, without hyphens ('-') and all lowercase. For example a9d4456e-349f-46f5-bc73-54a8d523e504 => a9d4456e349f46f5bc7354a8d523e504 => 8e63086f • The adler32 checksum without leading zeroes of the resource group name. For example rg-my-cool-single-rg => 4e23077c When you put these together, if the storage account is in East US 2 region in a resource group called "rg-my-cool-single-rg" and with a Subscription ID of a9d4456e-349f-46f5-bc73-54a8d523e504, the storage account name would be cbimgeu28e63086f4e23077c For the adler32 calculation, you can use the Online ADLER32 Hash Calculator. > **Note:** Make sure to remove all leading zeroes from the calculator's output. For example, if your resource group name is "cdppoc", convert it as follows: cdppoc => 089d027a => 89d027a

	Naming convention
Multiple resource groups created by Cloudera	Enter a name that is the concatenation of the following three elements: • "cbimg" • Region identifier: The starting letters of the region where the SA is. For example, "eu" for East US, or "eu2" for East US 2 • Subscription ID, without hyphens ('-') and all lowercase. For example, if your subscription ID is a9d4456e-349f-46f5-bc73-54a8d523e504, you should convert it to a9d4456e349f46f5bc7354a8d523e504 When you put these together, if the storage account is in East US 2 region in subscription a9d4456e-349f-46f5-bc73-54a8d523e504, the storage account name would be cbimgeu2a9d4456e349f46f5bc7354a8d523e504

b. Disable hierarchical namespace. Make sure not to omit this setting as it cannot be changed after the storage account is created.

Copying an image manually

In some cases where your organization's security policies require it, you must manually copy images to the storage account designated for image storage.

⚠️ **Caution:**

Cloudera creates new images for Cloudera frequently for all new Runtime versions and occasional OS bug fixes. Whenever a new image is available, it is not possible to launch new clusters until the image is copied. Furthermore, many features may not work correctly unless new images are copied to your storage account. For example, OS upgrade requires new images being present in your storage account.

The presence of new images can be monitored in the Cloudera image catalog at https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-cb-image-catalog.json or in Cloudera web interface > Cloudera Management Console> Shared Resources > Image Catalogs > cdp-default.

When copying the images, follow these high-level steps:

1. Create a container named "images" and copy all images to that container. Make sure to create a page blob.

   • Note the URL of the created container. You will need it to perform the copying process.
2. Identify the FreeIPA image that needs to be copied. You should use the latest image from the FreIPA image catalog available at https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-freeipa-image-catalog.json.
3. Identify the Data Lake and Cloudera Data Hub image(s) that need to be copied and note the VHD URLs. Both Data Lake and Cloudera Data Hub use the same image, so there is one image used for both for each Runtime version. There are two ways to find these images in the Cloudera image catalog:

   • In Cloudera web interface > Cloudera Management Console > Shared Resources > Image Catalogs > cdp-default > search for your Cloudera Runtime version > Click on UUID of your selected entry (For each Cloudera Runtime version, there should be only one entry for Azure cloud platform) > You will see VHD URLs, one per region.
   • At https://cloudbreak-imagecatalog.s3.amazonaws.com/v3-prod-cb-image-catalog.json.
4. Use azcopy tool to copy the images. When copying the images via the azcopy tool, you have two options for authentication and authorization:

   • Azcopy login: Perform azcopy login with a service principal that has the permission to copy into images container.
   • SAS token: Create a SAS token for the images container.

Option 1: Azcopy login

Issue the following command and it will perform an interactive login:

```
azcopy login
```

If you want to set up an automatic script then you can use:

```
azcopy login --service-principal \
  --application-id <YOUR_APPLICATION_ID> \
  --tenant-id <YOUR_TENANT_ID>
```

Note:

On some distributions (such as Ubuntu Linux) login executes fine, but logout fails (and so would any further authorization). For troubleshooting steps, refer to https://github.com/Azure/azure-storage-azcopy/issues/452.

To run azcopy, use:

```
azcopy copy \
  'https://<SOURCE_BLOB_URL>' \
  'https://<DESTINATION_BLUB_URL>
```

As mentioned earlier, the https://<SOURCE_BLOB_URL> can be found in the image catalog.

Option 2: SAS-token

Go to the Azure portal and generate a SAS-token. Then, you can issue azcopy:

```
azcopy copy \
  'https://<SOURCE_BLOB_URL>' \
  'https://<DESTINATION_BLUB_URL>?<YOUR_SAS_TOKEN>'
```

As mentioned earlier, the https://<SOURCE_BLOB_URL> can be found in the image catalog.

For more information about Azure copy syntax and examples, refer to Get started with AzCopy in Azure docs.

Creating an image resource

After the copying process is complete, create an image resource from each copied VHD.

In Azure Portal you can do that as follows:

1. Navigate to Images and click on Add.

2. Provide the following information:

Home > Images >

Create an image

Name *

freeipa-cdh--2010061458.vhd-westeurope

Subscription *

azure-se-cdp-sandbox-env

Resource group *

steffen-cdp-single-rg

Create new

Location *

(Europe) West Europe

Zone resiliency

On | Off

OS disk

OS type *

Windows | Linux

VM generation *

Gen 1 | Gen 2

Storage blob *

https://cbimgwe9fd109a759c70805.blob.core.windows.net/images/freeipa-cdh--2010061458.vhd | Browse

Storage type *

Standard HDD

Host caching *

Read/write

Data disks

Create | Automation options

3. Once done, click Create.

Supported Azure block storage

Review the list of supported Azure disk types for Cloudera.

Cloudera supports the following Azure disk types:

Cloudera name	Azure name
Standard_LRS	Standard HDD
StandardSSD_LRS	Standard SSD
Premium_LRS	Premium SSD

Related Information
Azure managed disk types

Azure Database for PostgreSQL

Cloudera provisions a PostgreSQL database as part of the environment creation process (for Data Lake). Prior to creating an Azure environment in Cloudera, the Azure Database for PostgreSQL must be enabled in the regions that you want to use with Cloudera.

Furthermore, Cloudera Data Warehouse service provisions PostgreSQL databases for its data warehouses.

Enabling Azure Database for PostgreSQL

The Azure Database for PostgreSQL service is often not enabled by default in every region, and your Azure administrator may need to contact Azure Support to have it enabled. For information on submitting this request, refer to Azure region access request process documentation. When making this request, do the following:

- Specify "Azure Database for PostgreSQL" in the Subscription field in Step 1.
- Specify "Azure Database for PostgreSQL" in the Region to Enable field in Step 4.
- Finally, in Step 4, specify 100 in the Planned Compute Usage in Cores field. This limits the pool of VCores per region to 100 VCores, limiting the total number of PostgreSQL servers that Cloudera can provision.

Root certificates for Azure Database

Cloudera products use the following root certificates in Azure Database for PostgreSQL SSL/TLS communication:

- BaltimoreCyberTrustRoot
- DigiCertGlobalRootG2
- DigiCertGlobalRootCA
- MicrosoftRSARootCertificateAuthority2017

This means that any update or rotation for the intermediate certificate chain does not have an affect on the root certificates used in Cloudera environments or their integrations.

Cluster deployments with Azure Policies for PostgreSQL

Cluster deployments with Azure Policies for PostgreSQL are not supported. Ensure that none of the Azure Policy built-in definitions for Azure Database for PostgreSQL are turned on or enforced. This is necessary because some policies (such as log_duration) could cause such performance degradation, making your clusters practically unusable.

Related Information
Private endpoint for Azure Postgres

Encrypting Azure resources with customer managed keys

By default, local Data Lake, FreeIPA, AKS Volumes, AKS Secrets, and Cloudera Data Hub disks attached to Azure VMs and the PostgreSQL server instance used by the Data Lake and Cloudera Data Hub clusters are encrypted with server-side encryption (SSE) using Platform Managed Keys (PMK), but you can optionally configure SSE with Customer Managed Keys (CMK).

The CMK can be specified during environment registration and, if present, is used for encrypting Data Lake, FreeIPA, AKS Volumes, AKS Secrets, and Cloudera Data Hub disks and PostgreSQL server instances.

The disks that are attached to the VMs of the Data Lake, FreeIPA, and Cloudera Data Hub clusters will be associated with a Disk Encryption Set (DES) that is created with the key URL as the underlying encryption key version. This disk encryption set is also used for volume encryption for AKS clusters. The DES dedicated to the Cloudera environment will be created in the resource group of the environment before the FreeIPA launch at the beginning of the environment creation process.

> **Note:** You must configure a managed identity to use CMK for AKS secrets. AKS CMK secret encryption will access the key vault directly using the Konnectivity agent and this managed identity. For more information, see the Using a user-managed identity for AKS clusters documentation.
>
> There is a limitation with the Konnectivity agent in Azure that requires the Key Vault firewall to allow public access. **This prevents a fully private setup.** For more infromation, see the Limitations in the KMS etcd encryption with AKS documentation.

When meeting Azure requirements for Cloudera, you should do the following:

- Add additional permissions for Cloudera provisioning credential
- Create a key vault and a vault key

- If you are using Azure Database for PostgreSQL Flexible Server with Cloudera, you can optionally use the CMK used for encrypting VM disks for encrypting the Azure Flexible Server database instance used by Cloudera. In this case, you should create a managed identity.

  Note:

  The CMK used for VM encryption can also be used for encrypting Azure Database for PostgreSQL Flexible Server used by Cloudera; this can be done through a user-supplied managed identity, as described in Configuring a CMK for data encryption in Azure Database for PostgreSQL Flexible Server and Managed identity for encrypting Azure Database for PostgreSQL Flexible Server. If the Flexible Server encryption with a CMK is set up, the permissions for accessing the CMK in Key Vault are conveyed by the user-supplied managed identity. In this case, the Key Vault resource may be configured with the "access policy" model or "RBAC" access control model. Cloudera will not make any changes to the Key Vault access policy; all permissions for accessing the Key Vault are encapsulated by the user-supplied managed identity. When using the "access policy" model, granting prior access for the user-supplied managed identity in the Key Vault access policy is your responsibility.

  In other words, when you provide the managed identity for CMK, VM disk encryption (via the DES resource) works identical to Flexible Server encryption. **If the managed identity for encryption is configured, it will also be used for enabling CMK secret encryption on AKS clusters.** This comes with additional requirements that is described in Using a user-managed identity for AKS clusters.

  If the user-supplied managed identity is not provided, the permissions for accessing the CMEK in Key Vault are conveyed using a system-assigned managed identity created by Cloudera. In this case, the Key Vault resource must be configured with the "access policy" model. Cloudera will automatically grant and revoke permissions for the DES system-assigned managed identity principal in the Key Vault resource access policy.

  For more information about Azure Key Vault's authorization systems, see Azure role-based access control (Azure RBAC) vs. access policies (legacy).

Add additional permissions to your Azure policy

Make sure that the following additional permissions are set up for the Azure credential used in Cloudera environment creation, in addition to what is documented in Azure permissions.

All of them are actions and shall be granted at the scope of the resource group hosting the Cloudera environment:

```
"Microsoft.KeyVault/vaults/read",
"Microsoft.KeyVault/vaults/write",
"Microsoft.KeyVault/vaults/deploy/action",
"Microsoft.Compute/diskEncryptionSets/read",
"Microsoft.Compute/diskEncryptionSets/write",
"Microsoft.Compute/diskEncryptionSets/delete",
"Microsoft.DBforPostgreSQL/servers/read",
"Microsoft.DBforPostgreSQL/servers/keys/write",
"Microsoft.KeyVault/vaults/accessPolicies/write"
```

Note:

When a user-supplied managed identity is created for Flexible Server encryption, the following permissions are not needed:

```
"Microsoft.KeyVault/vaults/accessPolicies/write"
"Microsoft.KeyVault/vaults/write"
```

See Managed identity for encrypting Azure Database for PostgreSQL Flexible Server.

The following table explains why Cloudera needs these permissions:

Permission	Description
Microsoft.KeyVault/vaults/read and Microsoft.KeyVault/vaults/write	Microsoft.KeyVault/vaults/read is required to read the vaults. Without this, vaults in your subscription cannot be detected by Cloudera. Specifically, Cloudera must update the Key Vault access policy to add an entry for the DES SP. For this, Cloudera invokes an "update key vault" operation, which Azure implements as the following series of steps: 1. Read Key Vault details, including its original access policy (Requires Microsoft.KeyVault/vaults/read). 2. Add entry for DES SP. 3. Write Key Vault details, including the updated access policy. In this step, Microsoft.KeyVault/vaults/write is required for adding an entry for the DES SP for the operations get, unwrap key and wrap key.
Microsoft.KeyVault/vaults/write	This is required to update the access policies for the DES created in the vault.
Microsoft.KeyVault/vaults/deploy/action	This is required to create DES resources. Specifically, it's required for DES SP creation (performed automatically alongside the DES creation) that will ultimately be used to access the Key Vault from the DES.
Microsoft.Compute/diskEncryptionSets/read	This is required to check for the existence of and fetch the properties and status of DES resources created by Cloudera.
Microsoft.Compute/diskEncryptionSets/write	This is required to create DES resources.
Microsoft.Compute/diskEncryptionSets/delete	This is required to delete the DES during environment termination.
Microsoft.DBforPostgreSQL/servers/keys/write and Microsoft.KeyVault/vaults/accessPolicies/write	This is required for setting up access policies and keys via an Azure Resource Manager template.

Create a vault and add a vault key

You can use your existing vault and vault key or create a new vault and vault key.

Regardless of which of the two options you choose, the vault and the vault key must fulfill the following requirements:

- The key vault must have purge protection enabled and be located in the same subscription and region as the target Cloudera environment.
- The CMK must be an RSA key with a size of 2048 bits.

> **Note:** The CMK must be an RSA key with a size of 2048 bits as Cloudera uses same key for all the resources and sizes other than 2048 bits are not supported by Azure for PostgreSQL.

- The number of Disk Encryption Set (DES) resources is limited to 1000 per region per subscription. In the present implementation, a single DES is created for each Cloudera environment, so this permits at most 1000 environments created in that region/subscription. The actual practical limit may be lower due to the limits set for other resource types.

You should also review the Azure-imposed restrictions for CMKs used for disk encryption, described in Server-side encryption of Azure Disk Storage: Restrictions in Azure documentation.

Using an existing vault and vault key

If you have an existing key vault in Key Vaults in your Azure Portal:

1. Navigate to the key vault's Overview page and verify that the following parameters are set to the correct values:

   - The Region matches the target region of the Cloudera environment. Storage accounts can be in different resource groups than the key vault, provided the location/region is the same.
   - The Purge protection is enabled.

2. Next, navigate to the vault key and:

   - Make sure that it is an RSA key with a size of 2048 bits.
   - Copy the key identifier (which is a HTTPS URL) for the key that is created. You will need to provide it during Cloudera environment registration later.

Creating a new vault and vault key

To create a vault and vault key, perform the following steps on your Azure portal:

1. Create a key vault in the same region and resource group as the one that you would like to use for registering the Cloudera environment.

   To create a key vault, navigate to Key vaults in Azure Portal and click on +New or on Create key vault. When providing key vault parameters, make sure that:

   - The Region matches the target region of the Cloudera environment. Storage accounts can be in different resource groups than the key vault, provided the location/region is the same.
   - The Purge protection is enabled.
   - Provide other parameters based on your organization's requirements. For instructions, see Create a vault in Azure documentation. Once done, click Create.

   The following screenshot points out these three important parameters when creating a new key vault:

2. Generate or import a key in the previously created key vault. Make sure that it is an RSA key with a size of 2048, 3072 or 4096 bits.

To generate a key, on your key vault's properties pages, select Keys and then click on Generate/Import:

Next provide key name and key type (make sure to select RSA). For detailed instructions, see Add a key to Key Vault in Azure documentation.

3. Once the key has been created, navigate to the vault key details and copy the key identifier (which is a HTTPS URL) for the key that is created. You will need to provide it during Cloudera environment registration later.

4. If you need encryption on the Azure storage account, you can set it up using Azure portal with the same or different encryption key. You can do this from the Encryption section of your storage account settings.

> **Note:** The key used for Azure storage encryption can be the same or different from the key vault used for disk encryption. Note that the storage account can be in a different resource group, but should be in the same location/region and subscription as the key vault.

Managed identity for encrypting Azure Database for PostgreSQL Flexible Server

If you are using Azure Database for PostgreSQL Flexible Server with Cloudera, you can optionally use the CMK used for encrypting VM disks for encrypting the Azure Flexible Server database instance used by Cloudera.

> **Note: Configuring this managed identity for Azure Database for PostgreSQL Flexible Server will also enable the CMK for AKS secrets.** This comes with additional requirements that are described in Using a user-managed identity for AKS clusters documentation.
>
> AKS CMK secret encryption will access the key vault directly using the Konnectivity agent. There is a limitation with the Konnectivity agent in Azure that requires the Key Vault firewall to allow public access. **This prevents a fully private setup**. For more information, see the Limitations in the KMS etcd encryption with AKS documentation.

In addition to the CMEK-related permissions described earlier, you should create a managed identity in Azure and assign to it an appropriate role with permissions that allow accessing the target key vault of the CMEK for cryptographic operations needed for storage encryption using that CMEK. In particular, the following dataActions shall be granted to the role:

```
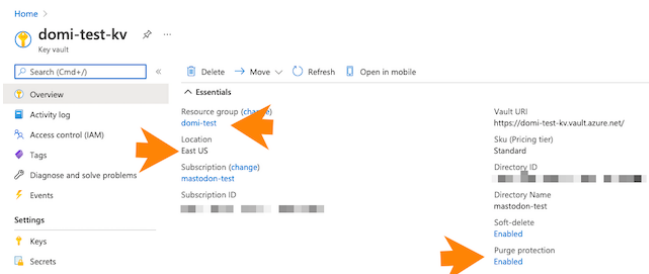"dataActions": [
"Microsoft.KeyVault/vaults/keys/read",
"Microsoft.KeyVault/vaults/keys/wrap/action",
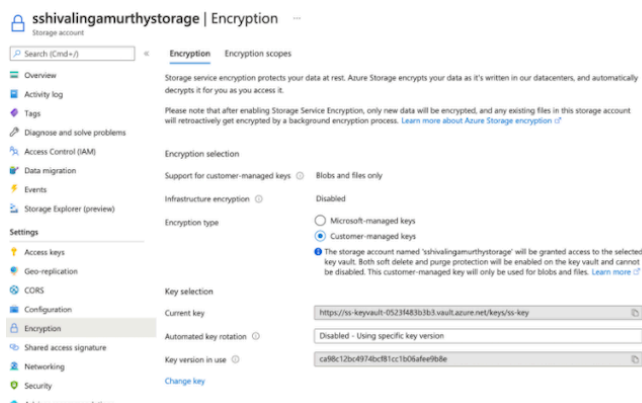"Microsoft.KeyVault/vaults/keys/unwrap/action"
],
```

There are two options:

- Using a built-in role provided by Azure:

  Key Vault Crypto Service Encryption User contains the bare minimum set of cryptographic permissions, although the Microsoft.EventGrid/* permissions are actually not needed. Alternatively, Key Vault Crypto User is also satisfactory, although it grants too many extra permissions on top of the ones that are strictly required.

- Using a custom role:

  We suggest cloning the Key Vault Crypto Service Encryption User and possibly removing the Microsoft.EventG rid/* permissions, which are not needed. Note that all the DataActions for Microsoft.KeyVault/* specified in this role are needed.

In either of the two cases, the role assignment of the managed identity should be scoped at the target Key Vault of the CMEK.

Note: If the managed identity for CMEK is absent during environment creation but the CMEK key URL is provided, the Disk Encryption Set (DES) resource is created using a system-assigned managed identity.

For information about providing the managed identity in Cloudera, see Configuring a CMK for data encryption in Azure Database for PostgreSQL Flexible Server.

Encrypting a storage account with a key vault that has role-based access control

To encrypt an ADLS Gen2 storage account that you would like to use with Cloudera with a key vault which has role-based access control set up, you need to perform the following steps on Azure Portal.

There are two scenarios described below:

- The first scenario assumes that you have not yet created the ADLS Gen2 storage account required for Cloudera on Azure.
- The second scenario assumes that you have already created the ADLS Gen2 storage account storage account required for Cloudera on Azure.

New storage account

The following steps should be performed in addition to the usual steps while creating the ADLS Gen2 account that you are planning to use with Cloudera.

Steps

1. Create a managed identity. Let's call it "key-vault-rbac".
2. Create a key vault.

   a. The key vault should have "purge protection" and "soft-delete" enabled.
   b. The key vault should be located in the same subscription and region as the target Cloudera environment.
   c. Set up the key's access policy to "Azure role-based access control".
3. Navigate to the access control for this key vault and:

   a. Click on Add role assignment.
   b. Assign the "Key Vault Administrator" role to all of the following users:

      - The user who created this key vault
      - The user(s) who register a Cloudera environment using this key vault
      - All the users/managed Identities who will be accessing this key vault

        Note: Without this step the key vault will not be accessible.

4. Once the key vault is created, create an RSA key with the size of 2048 bits.
5. Navigate to the access control for this key vault and:

   a. Click on Add role assignment.
   b. Assign the "Key Vault Crypto Service Encryption User" role to the "key-vault-rbac" managed identity created earlier. This will enable access to this key vault from the storage account.

6. Create the storage account and the managed Identities as mentioned in Minimal setup for cloud storage. During storage account creation, choose the following options to enable the customer managed key encryption for the storage account:

    a. Set Enable support for customer-managed keys to "All service types".

> **Note:** This is a mandatory step as this setting cannot be added or changed later.

    b. Set Identity type to "User-assigned".

    c. Set User-assigned identity to the "key-vault-rbac" managed identity created earlier.

Existing storage account

In case your ADLA Gen2 storage account already exists, perform the following steps instead of the ones above. The requirement is that the storage account must have been created with "Enable support for customer-managed keys" set to "All service types". This cannot be set once the storage account exists, so if the storage account does not have this set, you cannot use it for this use case.

Steps

1. Create a managed identity. Let's call it "key-vault-rbac".

2. Navigate to the access control for the key vault that you are using for Cloudera and:

    a. Click on Add role assignment.

    b. Assign the "Key Vault Crypto Service Encryption User" role to the "key-vault-rbac" managed identity created earlier. This will enable access to this key vault from the storage account.

3. To enable the customer managed key encryption for the storage account used in Cloudera, the following options must be chosen during storage account creation:

    a. Verify that Enable support for customer-managed keys is set to "All service types".

> **Note:** This option can only be set during storage account creation and cannot be changed later.

    b. Set Identity type to "User-assigned".

    c. Set User-assigned identity to the "key-vault-rbac" managed identity created earlier.

Azure Files storage account and file share for Cloudera AI

If you would like to use the Cloudera AI data service, create Azure Files storage account and file share. Azure Files NFS v4.1 is a managed, POSIX compliant NFS service on Azure. The file share is used to store files for the Cloudera AI infrastructure and Cloudera AI workbenches.

See Create Azure Files Storage Account and File Share.

Azure Files NFS for Cloudera AI

If you would like to use the Cloudera AI data service, Azure Files NFS is the recommended NFS service.

If you would like to use Cloudera AI, you should create an Azure Files storage account and file share to store files for the Cloudera AI infrastructure and Cloudera AI workbenches, as described in Create Azure Files Storage Account and File Share.

> **Note:** As an alternative to Azure Files NFS, you can configure an NFS server that is external to the Cloudera AI cluster. This is not the recommended approach. See Other NFS Options.

Azure quota limits

When you create your Azure Portal subscription, Azure Portal sets limits to the resources available to you. The limits can vary by region.

To register an Azure environment in Cloudera, you may need to increase some of these limits for the region(s) that you are planning to use. Cloudera creates resources such as VMs in your Azure subscription. For example:

- If you are planning to use the Cloudera AI service, review Limitations on Azure in Cloudera AI docs.
- Depending on the number of clusters that Cloudera creates in your Azure subscription, you might need to raise the limits for certain resources such as VMs and vCPUs in your Azure subscription. For a complete list of Azure resources used by Cloudera refer to Azure resources used by Cloudera.

If you require more resources than the limit set by Azure, you can create a support request on your Azure Portal. For information on Azure quotas, refer to Azure subscription and service limits, quotas, and constraints.

Overview of Azure resources used by Cloudera

The following Azure resources are used by Cloudera and Cloudera services.

Azure resources created for a Cloudera environment

When a Cloudera environment is created, a FreeIPA cluster and a Data Lake cluster are created.

The following Azure resources are created for FreeIPA (one per environment):

Resource	Description	Naming convention
Virtual Private Network (VNet)	If during environment creation you select to have a new VNet and subnets created, then a new VNet and subnets are created on your Azure account. Alternatively, you can provide your own existing VNet and subnets. In both cases (new and existing VPC), all compute resources that Cloudera provisions for the environment and Cloudera services are provisioned into the VNet specified during environment creation.	Specified by customer
Resource group for FreeIPA resources	If you chose for Cloudera to create multiple resource groups, a resource group is created to group all the resources created for FreeIPA. This resource group is not created if you chose to use a single existing resource group.	<env-name>-freeipa-<numeric-id>
Virtual Machines (VMs)	During environment creation, two or three Standard_DS3_v2 VMs are provisioned for the FreeIPA HA server by default. The number of VMs depends on the selected Data Lake type.	<env-name>-freeipa-<numeric-id>m0
OS disk	An OS disk is provisioned for the FreeIPA VM.	<env-name>-freeipa-<numeric-id>m0
Network interface	One network interface card (NIC) is provisioned for the FreeIPA VM.	<env-name>-freeipa-<numeric-id>m0
Public IP address	If you choose to use public IPs, your VM is assigned a public IP address.	<env-name>-freeipa-<numeric-id>m0

Resource	Description	Naming convention
Network security group	Network security groups define inbound and outbound access to the instances.If during environment creation you choose to have new security groups created, then they are created on your Azure account. Alternatively, you can provide your own existing security groups.	master0-<env-name>freeipa<numeric-id>sg
ADLS Gen2 storage account for storing operating system images	By default, Cloudera creates an ADLS Gen2 storage account that is used solely for storing operating system images. If required, you can optionally pre-create this account and copy the required images.	The name is the concatenation of the following three elements: • "cbimg" • Region identifier: The starting letters of the region where the SA is. For example, "eu" for East US, or "eu2" for East US 2. • Subscription ID, without hyphens ('-') and all lowercase. For example, if your subscription ID is a9d4456e-349f-46f5-bc73-54a8d523e504,you should convert it to a9d4456e349f46f5bc7354a8d523e504. For example, the name of a storage account in East US 2 region in subscription a9d4456e-349f-46f5-bc73-54a8d523e504 would be cbimgeu2a9d4456e349f46f5bc7354a8d523e504.
Resource group for the ADLS account used for storing operating system images	If you chose for Cloudera to create multiple resource groups, a separate resource group is created for the ADLS Gen2 account mentioned above. **Note:** To speed up future environment deployment, this resource group are its content are not deleted during environment deletion. This resource group is not created if you chose to use a single existing resource group.	cloudbreak-images

In addition, the following resources are created for each Data Lake (one per environment):

Resource	Description	Naming convention
Resource group	If you chose for Cloudera to create multiple resource groups, two resource groups are created: One new resource group is created to group all the resources created for the Data Lake and another resource group is created for the database used by the Data Lake. These resource groups are not created if you choose to use a single existing resource group.	<dl-name><numeric-id>

Resource	Description	Naming convention
Virtual Machines (VMs)	VMs are provisioned for the Data Lake nodes: **Enterprise:** • Two Standard_D4s_v5 instances (core) • Two Standard_D8s_v5 instances (gateway) • Three Standard_D8s_v5 instances (core) • One Standard_D4s_v5 instance (auxiliary) • Two Standard_D2s_v5 instances (idBroker) **Light duty:** • One Standard_E8s_v5 instance (Data Lake Master node) • One Standard_D2s_v5 instance (IDBroker). **Medium duty:** • One Standard_D4s_v5 instance (core) • Two Standard_D8s_v5 instances (gateway) • Three Standard_D8s_v5 instances (core) • One Standard_D4s_v5 instance (auxiliary) • Two Standard_D2s_v5 instances (idBroker) **Micro duty:** • One Standard_D2s_v5 instance (idBroker) • One Standard_D8s_v5 instance (gateway)	<dl-name><numeric-id>[m0|i1]

Resource	Description	Naming convention
OS disks	An OS disk is provisioned for each VM.	<env-name>-freeipa-<numeric-id>m0
Attached disks	An attached disk (StandardSSD_LRS) is provisioned for each VM.	<dl-name><numeric-id>-[m0|i1]<index>-<timestamp>
Network interface	One network interface card (NIC) is provisioned for each VM.	<dl-name><numeric-id>[m0|i1]
Public IP address	If you choose to use public IPs, each of the VMs is assigned a public IP address.	<dl-name><numeric-id>[m0|i1]
Network security groups	Network security groups define inbound and outbound access to the instances.If during environment creation you choose to have new security groups created, then they are created on your Azure account.	[master|idbroker]<dl-name><numeric-id>sg
Availability set	One availability set is created for the master host group only.	<dl-name>-master-as
Resource group for external DB	If you chose for Cloudera to create multiple resource groups, a resource group is created for the external database. This resource group is not created if you chooe to use a single existing resource group.	<env-name>-dbstck-<numeric-id>
Azure Database for PostgreSQL server	An RDS instance is provisioned for Cloudera Manager, Ranger, and Hive MetaStore. When creating Flexible Server, Cloudera automatically chooses the latest generation of Standard_E4s instance family that is supported in the given region, for example, Standard_E4ds_v5, Standard_E4ds_v4 or Standard_E4s_v3 with 128 GB of storage. For more information, see Azure regions. If you choose to use Single Server, a database instance MO_Gen5_4 with 100 GB of storage) is provisioned.	dbsrv-<numeric-id>
ADLS Gen2 storage	Prior to registering your environment in Cloudera, you should create ADLS Gen2 storage containers as instructed in Cloudera documentation.	Specified by customer
Managed identities	Prior to registering your environment in Cloudera, you should create managed identities as instructed in Cloudera documentation.	Specified by customer

Azure resources created for Cloudera Data Hub

The following Azure resources are created for the Cloudera Data Hub service:

Resource	Description	Naming convention
Resource group	If you chose for Cloudera to create multiple resource groups, for each Cloudera Data Hub cluster, a new resource group is created to group all the resources created for the cluster. This resource group is not created if you chose to use a single existing resource group.	<dh-name><numeric-id>
Virtual Machines (VMs)	A VM is created for each cluster node. The VM type varies depending on what you selected during Cloudera Data Hub cluster creation. For a list of supported VM types, refer to Cloudera on cloud service rates.	<dh-name><numeric-id><hostgroup abbr.><node index>

Resource	Description	Naming convention
OS disk	An OS disk is provisioned for each VM.	<dh-name><numeric-id>-osDisk<hostgroup abbr.><node index>
Attached Disks	An attached disk is provisioned for each VM, as specified during Cloudera Data Hub cluster creation. The disk size is selected during cluster creation.	<dh-name><numeric-id>-<hostgroup abbr.><node index>-<disk counter>-<timestamp>
Network interface	One network interface card (NIC) is provisioned for each VM.	<dh-name><numeric-id><hostgroup abbr.><node index>
Public IP address	If you choose to use public IPs, each of the VMs is assigned a public IP address.	<dh-name><numeric-id><hostgroup abbr.><node index>
Network security group	Network security groups define inbound and outbound access to the instances. If during environment creation you choose to have new security groups created, then they are created on your Azure account.	<hostgroup nm><dh-name><numeric-id>sg
Availability set	If the "Hardware and Storage" Advanced Options were used, one availability set is created for each host group. Otherwise, one availability set is created only for the host groups that contain KNOX and/or OOZIE service.	<dh-name>-<hostgroup>-as

Azure resources created for Cloudera Data Warehouse

The following Azure resources are created for the Cloudera Data Warehouse service:

Resource	Description
Resource Group	If you chose for Cloudera to create multiple resource groups, one resource group is created with the naming convention "<environment-id>-dwx-rg". This resource group is not created if you chose to use a single existing resource group.
Azure Kubernetes Service (AKS)	Cloudera creates an AKS cluster for each activated Cloudera Data Warehouse environment to host Kubernetes-based resources. The underlying compute, network resources are managed by Azure, including: • Resource group • Virtual machine scale sets • Load balancer(s) • Public IP address(es) • Network security group • Disk(s) For a list of supported VM types, refer to Cloudera on cloud service rates.
Azure Database for PostgreSQL server	PostgreSQL database (General Purpose, Gen5, 4 vCore) is created for Cloudera Data Warehouse to store configuration data.

Azure resources created for Cloudera AI

The following Azure resources are created for the Cloudera AI service:

Resource	Description
Resource groups	If you chose for Cloudera to create multiple resource groups, one resource group is created with the naming convention "liftie-<unique string>" (which has an AKS cluster of the same name). This resource group is not created if you chose to use a single existing resource group.

Resource	Description
Azure Kubernetes Service (AKS)	Cloudera creates an AKS cluster for each Cloudera AI workbench to host Kubernetes-based resources. The underlying compute, network resources are managed by Azure, including: • Resource group • Virtual machine scale sets • Load balancer(s) • Public IP address(es) • Route table • Network security group • Azure disk(s) (Premium_LRS) For a list of supported VM types, refer to Cloudera on cloud service rates.
Log analytics workspace	A logs analytics workspace is created for storing log data.
Azure Files Storage account	If you choose Azure Files NFS, you will need an existing Azure Files Storage account.

Azure resources created for Cloudera DataFlow

The following Azure resources are created for the Cloudera DataFlow service:

Resource	Description
Resource groups	If you chose for Cloudera to create multiple resource groups, one resource group is created with the naming convention "liftie-<unique string>" (which has an AKS cluster of the same name). This resource group is not created if you chose to use a single existing resource group.
Azure Kubernetes Service (AKS)	Cloudera creates an AKS cluster for the Cloudera DataFlow service. The underlying compute, network resources are managed by Azure, including: • Load balancer: Azure Load Balancer • Network security group • Public IP address(es) • Route table • Virtual machine scale set For a list of supported VM types, refer to Cloudera on cloud service rates.
Log analytics workspace	A logs analytics workspace is created for storing log data.
Azure Database for PostgreSQL	Azure Database for PostgreSQL is used for storing job-related metadata and histories.

Azure resources created for Cloudera Data Engineering

The following Azure resources are created for the Cloudera Data Engineering service:

Resource	Description
Resource groups	If you chose for Cloudera to create multiple resource groups, one resource group is created with the naming convention "liftie-<unique string>" (which has an AKS cluster of the same name). This resource group is not created if you chose to use a single existing resource group.

Resource	Description
Azure Kubernetes Service (AKS)	Cloudera creates an AKS cluster for each Cloudera Data Engineering Service. The underlying compute, network resources are managed by Azure, including: • Load balancer: Azure Load Balancer • Network security group • Public IP address(es) • Route table • Virtual machine scale set For a list of supported VM types, refer to Cloudera on cloud service rates.
Log analytics workspace	A logs analytics workspace is created for storing log data.
Azure Files	Azure Files	Microsoft Azure contains job resources, application code, Apache Airflow DAG files and any other uploaded files.
Azure Database for MySQL Server	Azure Database for MySQL is used for storing job related metadata, histories.

Azure resources created for Cloudera Operational Database

The following Azure resources are created for the Cloudera Operational Database service:

Resource	Description
Resource Group	If you chose for Cloudera to create multiple resource groups, a resource group is created which contains all of the nodes that comprise the Cloudera Operational Database database. This resource group is not created if you chose to use a single existing resource group.
Virtual Machines (VMs)	A compute VM is created for each node in a Cloudera Operational Database database. The instance type and managed storage are automatically determined by Cloudera Operational Database. Azure network security groups are automatically configured as a part of environment creation to define inbound and outbound network access to the created instances.
ADLS Gen2 storage	This existing blob storage account that you provided for the Data Lake to use for workload data storage is automatically used by the Cloudera Operational Database database for storage of data.

Using a user-managed identity for AKS clusters

When you are creating a Cloudera environment, you can specify a user-managed identity to be used for management of Azure resources.

⚠️ **Important:** Configuring the user-managed identity is optional when using secret-based authentication, but required when using certificate-based authentication. For more information, see the Azure credential prerequisites documentation.

You have the following options to use a user-managed identity:

• Configure encryption on the environment and specify a managed identity
• Configure a specific user managed identity for AKS clusters

Encryption Managed Identity

It is possible to configure a managed identity when setting up the encryption for the Cloudera environment. This managed identity is primarily used for the encryption of the Azure Database for PostgreSQL Flexible Server, but can also serve as the AKS managed identity.

For more information about the Azure Database for PostgreSQL Flexible Server, see the Encrypting VM disks with customer managed keys documentation.

If managed identity for encryption is enabled, the managed identity is used for AKS as well, if you do not to provide a separate user-managed identity for AKS. The managed identity need to have the following permissions in addition to what is required for the PostgreSQL Flexible Server:

- Contributor on the resource group the AKS cluster is being created in
- Contributor on the resource group containing the network to be used
- Managed Identity Operator on the resource group containing additional managed identities (logger identity and so on)

You can use the identity create command in Azure CLI to create the managed identity and the role assignment create command to assign roles as shown in the following example:

```
$ az identity create -n AKS-IDENTITY-EXAMPLE -g TEST-RESOURCE-GROUP -
l WESTUS2
{
  "clientId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "id": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourcegro
ups/TEST-RESOURCE-GROUP/providers/Microsoft.ManagedIdentity/userAssignedIde
ntities/AKS-IDENTITY-EXAMPLE",
  "location": "westus2",
  "name": "TEST-AKS-IDENTITY-EXAMPLE",
  "principalId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "resourceGroup": "<resource-group-name>",
  "systemData": null,
  "tags": {},
  "tenantId": "49D9544F-09A2-4360-BFA7-7E08AD057157",
  "type": "Microsoft.ManagedIdentity/userAssignedIdentities"
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Contributor" --scope "/subscriptions/43BA24CE-DF3C-4308-901D-
A7006E345E55/resourceGroups/AKS-RESOURCE-GROUP"
{
  "condition": null,
  "conditionVersion": null,
  "createdBy": null,
  "createdOn": "2023-11-27T14:00:50.679610+00:00",
  "delegatedManagedIdentityResourceId": null,
  "description": null,
  "id": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourceGro
ups/TEST-RESOURCE-GROUP/providers/Microsoft.Authorization/roleAssignmen
ts/AAA3BEC0-BF57-4E09-8B85-8B1652F3647A",
  "name": "AAA3BEC0-BF57-4E09-8B85-8B1652F3647A",
  "principalId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "principalType": "ServicePrincipal",
  "resourceGroup": "TEST-RESOURCE-GROUP",
  "roleDefinitionId": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/
providers/Microsoft.Authorization/roleDefinitions/098C7804-2B99-49BD-9DBD-
EB188E4213C6",
  "scope": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourceGro
ups/AKS-RESOURCE-GROUP",
  "type": "Microsoft.Authorization/roleAssignments",
  "updatedBy": "BEA2645F-78EB-4CC6-A269-F6108A74D48D",
  "updatedOn": "2023-11-27T14:00:51.037481+00:00"
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Contributor" --scope "/subscriptions/43BA24CE-DF3C-4308-901D-
A7006E345E55/resourceGroups/NETWORK-RESOURCE-GROUP"
{
  ...
}
```

```
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Managed Identity Operator" --scope "/subscriptions/43BA24CE-
DF3C-4308-901D-A7006E345E55/resourceGroups/MANAGED-IDENTITY-RESOURCE-GROUP"
{
  ...
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Key Vault Crypto Officer" --scope "/subscriptions/43BA24CE-
DF3C-4308-901D-A7006E345E55/resourceGroups/MANAGED-IDENTITY-RESOURCE-GROUP/
providers/Microsoft.KeyVault/vaults/KEY-VAULT"
{
  ...
}
```

AKS Specific Managed Identity

You can provide a specific managed identity that is only used for the AKS clusters.

The managed identity need to have the following permissions:

- Contributor on the resource group the AKS cluster is being created in
- Contributor on the resource group containing the network to be used
- Managed Identity Operator on the resource group containing additional managed identities (logger identity etc.)

You can use the identity create command in Azure CLI to create the managed identity and the role assignment create command to assign roles as shown in the following example:

```
$ az identity create -n AKS-IDENTITY-EXAMPLE -g TEST-RESOURCE-GROUP -
l WESTUS2
{
  "clientId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "id": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourcegro
ups/TEST-RESOURCE-GROUP/providers/Microsoft.ManagedIdentity/userAssignedIde
ntities/AKS-IDENTITY-EXAMPLE",
  "location": "westus2",
  "name": "TEST-AKS-IDENTITY-EXAMPLE",
  "principalId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "resourceGroup": "<resource-group-name>",
  "systemData": null,
  "tags": {},
  "tenantId": "49D9544F-09A2-4360-BFA7-7E08AD057157",
  "type": "Microsoft.ManagedIdentity/userAssignedIdentities"
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Contributor" --scope "/subscriptions/43BA24CE-DF3C-4308-901D-
A7006E345E55/resourceGroups/AKS-RESOURCE-GROUP"
{
  "condition": null,
  "conditionVersion": null,
  "createdBy": null,
  "createdOn": "2023-11-27T14:00:50.679610+00:00",
  "delegatedManagedIdentityResourceId": null,
  "description": null,
  "id": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourceGro
ups/TEST-RESOURCE-GROUP/providers/Microsoft.Authorization/roleAssignmen
ts/AAA3BEC0-BF57-4E09-8B85-8B1652F3647A",
  "name": "AAA3BEC0-BF57-4E09-8B85-8B1652F3647A",
  "principalId": "25104331-3CDC-4A5E-9E8B-8AFC066C9BEB",
  "principalType": "ServicePrincipal",
  "resourceGroup": "TEST-RESOURCE-GROUP",
  "roleDefinitionId": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/
providers/Microsoft.Authorization/roleDefinitions/098C7804-2B99-49BD-9DBD-
EB188E4213C6",
```

```
   "scope": "/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourceGro
ups/AKS-RESOURCE-GROUP",
   "type": "Microsoft.Authorization/roleAssignments",
   "updatedBy": "BEA2645F-78EB-4CC6-A269-F6108A74D48D",
   "updatedOn": "2023-11-27T14:00:51.037481+00:00"
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Contributor" --scope "/subscriptions/43BA24CE-DF3C-4308-901D-
A7006E345E55/resourceGroups/NETWORK-RESOURCE-GROUP"
{
   ...
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Managed Identity Operator" --scope "/subscriptions/43BA24CE-
DF3C-4308-901D-A7006E345E55/resourceGroups/MANAGED-IDENTITY-RESOURCE-GROUP"
{
   ...
}
$ az role assignment create --assignee 25104331-3CDC-4A5E-9E8B-8AFC066C9BEB
 --role "Key Vault Crypto Officer" --scope "/subscriptions/43BA24CE-
DF3C-4308-901D-A7006E345E55/resourceGroups/MANAGED-IDENTITY-RESOURCE-GROUP/
providers/Microsoft.KeyVault/vaults/KEY-VAULT"
{
   ...
}
```

If you have the managed identity for encryption enabled and a managed identity for AKS configured, the AKS managed identity must have the following permissions in addition to the already configured ones:

- If Key Vault is setup with RBAC permissions:

    - Key Vault Crypto Officer on Key Vault containing CMK if using managed identities with encryption
- If Key Vault is setup with Access policies:

    - Key Management Operations - ALL
    - Cryptographic Operations - ALL

⚠️ **Important:** There is a limitation with the Konnectivity agent in Azure that requires the Key Vault firewall to allow public access. This prevents a fully private setup. For more infromation, see the Limitations in the KMS etcd encryption with AKS documentation.

You can create an environment with a managed identity by using the following command:

```
$ cdp environments create-azure-environment --environment-name "CERT-ENV-
EXAMPLE" --credential-name "CERT-EXAMPLE" --data-services "azure={sharedMana
gedIdentity=/subscriptions/43BA24CE-DF3C-4308-901D-A7006E345E55/resourcegro
ups/TEST-RESOURCE-GROUP/providers/Microsoft.ManagedIdentity/userAssignedIde
ntities/AKS-IDENTITY-EXAMPLE}"
```

You can also update an existing environment to add an AKS identity with the following command:

```
$ cdp environments update-data-service-resources --environment "CERT-
ENV-EXAMPLE" --data-services "azure={sharedManagedIdentity=/subscriptio
ns/43BA24CE-DF3C-4308-901D-A7006E345E55/resourcegroups/TEST-RESOURCE-GROUP/
providers/Microsoft.ManagedIdentity/userAssignedIdentities/AKS-IDENTITY-
EXAMPLE}"
```

Azure outbound network access destinations

If you have limited outbound internet access (for example due to using a firewall or proxy), review this content to learn which specific outbound destinations must be available in order to register a Cloudera environment.

We recommend hostname-based policies, as some of the destination services do not have static IP addresses. IP address details in CIDR notation have been provided where static IPs are in-use.

> **Warning:** On 30 September 2025, Microsoft will retire the default outbound access connectivity for virtual machines in Azure. You must configure an explicit outbound connectivity method. For more information, see Cloudera Customer Advisory-866 and Transitioning from Azure default outbound access documentation.

> **Note:**
>
> If the cloud provider network that you would like to use for registering a Cloudera environment uses a custom DNS server that does not allow name resolution for public domain, you should add all the domains listed in the below tables to the DNS forwarder for name resolution.

> **Note:**
>
> On Azure, it is possible to define network security groups (NSGs) based on IP addresses, but not based on hostnames. At the same time, the egress filtering requirements documented here contain not only static IP addresses but also hostnames, where the IP address may change over time. This means that it is not possible to perform egress filtering using NSGs. Instead, if you would like to add outbound network access to your allow list based on hostnames, you should use Azure firewall with FQDN filtering in network rules. For that you can use either Azure hosted DNS or a custom DNS, as explained in Use FQDN filtering in network rules. Note that interfaces configured via Azure Private Link don't need to go through a proxy.

The following list includes general destinations as well as Azure-specific destinations.

General endpoints

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Control Plane API	All services	US-based Control Plane: api.us-west-1.cdp.cloudera.com EU-based Control Plane: api.eu-1.cdp.cloudera.com AP-based Control Plane: api.ap-1.cdp.cloudera.com	HTTPS with Cloudera-generated access key	TCP/443	Cloudera Control Plane REST API.
Cloudera CCMv1 Persistent Control Plane connection	All services	*.ccm.cdp.cloudera.com 44.234.52.96/27	SSH public/ private key authentication	TCP/6000-6049	One connection per cluster configured; persistent
Cloudera CCMv2 Persistent Control Plane connection	All services	US-based Control Plane: *.v2.us-west-1.ccm.cdp.cloudera.com 35.80.24.128/27 EU-based Control Plane: *.v2.ccm.eu-1.cdp.cloudera.com 3.65.246.128/27 AP-based Control Plane: *.v2.ccm.ap-1.cdp.cloudera.com 3.26.127.64/27	HTTPS with mutual authentication	TCP/443	Multiple long-lived/persistent connections

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Cloudera Databus Telemetry, billing and metering data	All services	US-based Control Plane: dbusapi.us-west-1.sigma.altus.cloudera.com api.us-west-1.cdp.cloudera.com https://cloudera-dbus-prod.s3.amazonaws.com EU-based Control Plane: api.eu-1.cdp.cloudera.com https://mow-prod-eu-central-1-sigmadbus-dbus.s3.eu-central-1.amazonaws.com https://mow-prod-eu-central-1-sigmadbus-dbus.s3.amazonaws.com AP-based Control Plane: api.ap-1.cdp.cloudera.com https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.ap-southeast-2.amazonaws.com https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.amazonaws.com	HTTPS with Cloudera-generated access key for dbus HTTPS for S3	TCP/443	Regular interval for telemetry, billing, metering services, and used for Cloudera Observability if enabled. Larger payloads are sent to a Cloudera managed S3 bucket.
Cloudera Observability Metrics System metrics collection	All services	US-based Control Plane: *.api.monitoring.us-west-1.cdp.cloudera.com EU-based Control Plane: *.api.monitoring.eu-1.cdp.cloudera.com AP-based Control Plane: *.api.monitoring.ap-1.cdp.cloudera.com	HTTPS	TCP/443	New as of March 2024
Cloudera Manager parcels Software distribution	All services	archive.cloudera.com	HTTPS	TCP/443	Cloudera's public software repository. CDN backed service; IP range not predictable.
RPMs Cloudera RPMs for workload agents	All services	cloudera-service-delivery-cache.s3.amazonaws.com	HTTPS	TPC/443	RPM packages for some workload components

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Container Images Software Distribution	Cloudera Data Engineering Cloudera Data Warehouse Cloudera AI	container.repo.cloudera.com container.repository.cloudera.com prod-us-west-2-starport-layer-bucket.s3.us-west-2.amazonaws.com prod-us-west-2-starport-layer-bucket.s3.amazonaws.com s3-r-w.us-west-2.amazonaws.com *.execute-api.us-west-2.amazonaws.com prod-eu-west-1-starport-layer-bucket.s3.eu-west-1.amazonaws.com prod-eu-west-1-starport-layer-bucket.s3.amazonaws.com s3-r-w.eu-west-1.amazonaws.com *.execute-api.eu-west-1.amazonaws.com prod-ap-southeast-1-starport-layer-bucket.s3.ap-southeast-1.amazonaws.com prod-ap-southeast-1-starport-layer-bucket.s3.amazonaws.com s3-r-w.ap-southeast-1.amazonaws.com *.execute-api.ap-southeast-1.amazonaws.com	HTTPS	TCP/443	CDN-backed and AWS ECR-backed services; IP range not predictable. container.repo.cloudera.com uses ECR backend which requires S3 URLs. IP geolocation attempts to select closest API and ECR backend; clients may be directed to any of the destinations.
Flow Definitions Cloudera AWS bucket with flow definitions	Cloudera DataFlow	US-based Control Plane: s3.us-west-2.amazonaws.com/dfx-flow-artifacts.mow-prod.mow-prod.cloudera.com EU-based Control Plane: cldr-mow-prod-eu-central-1-dfx-flow-artifacts.s3.eu-central-1.amazonaws.com AP-based Control Plane: cldr-mow-prod-ap-southeast-2-dfx-flow-artifacts.s3.ap-southeast-2.amazonaws.com	HTTPS (one way) IAM authentication	TCP/443	Outbound internet access to S3 hosts is necessary on all cloud providers when using Cloudera DataFlow as the workload needs to query outbound to an S3 location to retrieve the flow definition when creating a deployment.
Public Signing Key Retrieval	Cloudera Data Engineering Cloudera DataFlow	US-based Control Plane: consoleauth.altus.cloudera.com console.us-west-1.cdp.cloudera.com EU-based Control Plane: console.eu-1.cdp.cloudera.com AP-based Control Plane: console.ap-1.cdp.cloudera.com	HTTPS	TCP/443	Required to allow authentication to Cloudera Data Engineering virtual Cluster using a Cloudera Access Key.

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
Control Plane IAM API	Cloudera AI	US-based Control Plane: iamapi.us-west-1.altus.cloudera.com console.us-west-1.cdp.cloudera.com EU-based Control Plane: console.eu-1.cdp.cloudera.com AP-based Control Plane: console.ap-1.cdp.cloudera.com	HTTPS	TCP/443	For connecting to the IAMAPI for fetching the entitlement details.
AMPs Cloudera Accelerators for Machine Learning Projects	Cloudera AI	https://raw.githubusercontent.com https://github.com	HTTPS	TCP/443	Files for AMPs are hosted on GitHub.
AI Agent Studios	Cloudera AI	huggingface.co github.com github.infra.cloudera.com nodejs.org iojs.org container.repo.cloudera.com	HTTPS	TCP/443	Required for AI Agent Studios
Cloudera Copilot	Cloudera AI	bedrock-runtime.<region>.amazonaws.com api.gradio.app	HTTPS	TCP/443	Required for Cloudera Copilot AMP when deploying an LLM on AWS Bedrock.
Learning Hub	Cloudera AI	https://github.com/cloudera/learning-hub-content	HTTPS	TCP/443	Access Learning Hub in air-gapped environments

Azure-specific endpoints

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
General Azure guidelines	All services	See Safelist the Azure portal URLs on your firewall or proxy server for Azure egress best practices.			
Azure Data Lake Storage Gen 2	All services	<STORAGE_ACCOUNT_NAME>.dfs.core.windows.net	HTTPS Azure authentication	TCP/443	Azure Storage VPC endpoint is required (Microsoft.Storage).Replace the <STORAGE_ACCOUNT_NAME> with an actual storage account name.
Azure Database for Postgres	All services	*.postgres.database.azure.com	JDBC / Postgres binary protocol	TCP/5432	Azure SQL VPC endpoint is required (Microsoft.Sql).

Description/ Usage	Cloudera service	Destination	Protocol and Authentication	IP Protocol/Port	Comments
ARM to manage User Assigned Managed Identities	All services	management.azure.com	HTTPS Azure authentication	TCP/443	This can be allowed by using the AzureResourceManager Azure service tag. Additionally IP addresses to whitelist are available to download.
Microsoft Log Analytics	All services	*.agentsvc.azure-automation.net *.ods.opinsights.azure.com *.oms.opinsights.azure.com *.blob.core.windows.net	HTTPS Azure authentication	TCP/443	Optional, but may cause issues with Azure approved images if blocked.
Azure Kubernetes Services (AKS)	Cloudera Data Engineering Cloudera DataFlow Cloudera Data Warehouse Cloudera AI	See Outbound network and FQDN rules for Azure Kubernetes Service (AKS) clusters. **Note:** Cloudera uses AKS and has the same requirements. Therefore, your setup must fulfill the Azure EKS requirements mentioned in the linked documentation (such as port 9000). If you skip this step, your deployment will fail.			
Azure Database for MySQL	Cloudera Data Engineering	*.mysql.database.azure.com	JDBC / Postgres binary protocol	TCP/3306	Azure Database for MySQL
Azure files	Cloudera Data Engineering	*.file.core.windows.net	SMB	TCP/445	What is Azure Files?
Azure Files NFS	Cloudera AI	*.file.core.windows.net	NFS	TCP/2049	Create an NFS Azure file share
Digicert CA Certificate	Cloudera Data Engineering Cloudera DataFlow	www.digicert.com cacerts.digicert.com	HTTPS Azure authentication	TCP/443	Fetching TLS CA for Azure MySQL DB secure connection

Access to workload UIs

If you have restricted DNS or networking setup, make sure that *.cloudera.site is resolvable from your network so that members of your organization can access workload UIs.

Cloudera workloads (including Data Lake) use subdomains under cloudera.site to host various UI endpoints (Cloudera Manager, Ranger, Knox, Hue and so on). Cloudera automatically provisions these endpoints whenever a Data Lake, Cloudera Data Hub or another type of workload (for example, Virtual Warehouse in Cloudera Data Warehouse) is created, and routing is set up so that you can access these endpoints from your network.

The subdomains are assigned under cloudera.site using the following convention:

```
<endpoint-name>.<env-truncated-name>.<customer-workload-subdomain>.<regional
-subdomain>.cloudera.site
```

Supported browsers

Cloudera validates and tests against the latest version and supports recent versions of the following browsers:

• Google Chrome
• Mozilla Firefox

> **Note:**
> • Mozilla Firefox is not supported by Cloudera Data Engineering.
> • Certain accessibility features in Cloudera DataFlow do not work in Mozilla Firefox.

• Safari
• Microsoft Edge

Other resources

While this document attempts to provide a complete overview of cloud provider requirements, there is additional documentation that you should review if planning to deploy Cloudera data services.

The following table includes links to documentation that you should review if planning to deploy the respective Cloudera data service:

Cloudera data service	Documentation links
Cloudera Data Engineering	Cloudera Data Engineering Azure prerequisites Cloudera Data Engineering cost management Cloudera Data Engineering performance management
Cloudera Data Warehouse	Virtual Warehouse sizing requirements for public cloud environments Virtual Warehouse IP address and cloud resource requirements for environments on cloud Managing costs in the public cloud environments for Cloudera Data Warehouse Azure environments requirements checklist
Cloudera DataFlow	Azure requirements for Cloudera DataFlow Azure load balancers in Cloudera DataFlow Cloudera DataFlow limitations on Azure
Cloudera AI	Azure requirements for Cloudera AI workbenches Cloudera AI limitations on Azure Network Planning for Cloudera AI on Azure
Cloudera Operational Database	Cloudera Operational Database cloud checklist Azure requirements for Cloudera Operational Database deployment

Cloudera CIDR

List of Cloudera Control Plane source IP addresses

In cases where the Cloudera Control Plane is communicating directly with resources in the Cloudera environment, the Cloudera Control Plane source IP address will be from the ranges below.

Security groups deployed/managed by Cloudera Control Plane will automatically include the required rules. These source IP ranges will be included if the Cloudera Control Plane requires direct inbound access to the resource.

Cloudera Control Plane Region	IP Ranges
us-west-1	35.80.24.128/27, 35.166.86.177/32, 52.36.110.208/32, 52.40.165.49/32
eu-1	3.65.246.128/27
ap-1	3.26.127.64/27

If the security groups (or other network access policies) in the network path are not managed by the Cloudera Control Plane, ensure that inbound access from the appropriate range(s) are permitted.