Secure an Azure Machine Learning workspace with virtual networks

Article • 09/30/2022 • 15 minutes to read

APPLIES TO: ✓ Azure CLI ml extension v2 (current) ✓ Python SDK azure-ai-ml v2 (current)

Select the version of Azure Machine Learning SDK/CLI extension you are using:

In this article, you learn how to secure an Azure Machine Learning workspace and its associated resources in a virtual network.

> **Tip**
>
> This article is part of a series on securing an Azure Machine Learning workflow. See the other articles in this series:
>
> - Virtual network overview
> - Secure the training environment
> - Secure the inference environment
> - Enable studio functionality
> - Use custom DNS
> - Use a firewall
> - API platform network isolation
>
> For a tutorial on creating a secure workspace, see Tutorial: Create a secure workspace or Tutorial: Create a secure workspace using a template.

In this article you learn how to enable the following workspaces resources in a virtual network:

- ✓ Azure Machine Learning workspace
- ✓ Azure Storage accounts
- ✓ Azure Key Vault
- ✓ Azure Container Registry

Prerequisites

- Read the Network security overview article to understand common virtual network scenarios and overall virtual network architecture.
- Read the Azure Machine Learning best practices for enterprise security article to learn about best practices.
- An existing virtual network and subnet to use with your compute resources.

> ⓘ **Important**
>
> We do not recommend using the 172.17.0.0/16 IP address range for your VNet. This is the default subnet range used by the Docker bridge network. Other ranges may also conflict depending on what you want to connect to the virtual network. For example, if you plan to connect your on premises network to the VNet, and your on-premises network also uses the 172.16.0.0/16 range. Ultimately, it is up to **you** to plan your network infrastructure.

- To deploy resources into a virtual network or subnet, your user account must have permissions to the following actions in Azure role-based access control (Azure RBAC):
  - "Microsoft.Network/virtualNetworks/join/action" on the virtual network resource.
  - "Microsoft.Network/virtualNetworks/subnets/join/action" on the subnet resource.

  For more information on Azure RBAC with networking, see the Networking built-in roles

Azure Container Registry

- Your Azure Container Registry must be Premium version. For more information on upgrading, see Changing SKUs.
- If your Azure Container Registry uses a **private endpoint**, it must be in the same *virtual network* as the storage account and compute targets used for training or inference. If it uses a **service endpoint**, it must be in the same *virtual network* and

subnet as the storage account and compute targets.

- Your Azure Machine Learning workspace must contain an Azure Machine Learning compute cluster.

Limitations

Azure storage account

- If you plan to use Azure Machine Learning studio and the storage account is also in the VNet, there are extra validation requirements:
  - If the storage account uses a **service endpoint**, the workspace private endpoint and storage service endpoint must be in the same subnet of the VNet.
  - If the storage account uses a **private endpoint**, the workspace private endpoint and storage private endpoint must be in the same VNet. In this case, they can be in different subnets.

Azure Container Instances

When your Azure Machine Learning workspace is configured with a private endpoint, deploying to Azure Container Instances in a VNet is not supported. Instead, consider using a Managed online endpoint with network isolation.

Azure Container Registry

When ACR is behind a virtual network, Azure Machine Learning can't use it to directly build Docker images. Instead, the compute cluster is used to build the images.

> ⓘ **Important**
>
> The compute cluster used to build Docker images needs to be able to access the package repositories that are used to train and deploy your models. You may need to add network security rules that allow access to public repos, **use private Python packages**, or use **custom Docker images** that already include the packages.

> ⚠ **Warning**

If your Azure Container Registry uses a private endpoint or service endpoint to communicate with the virtual network, you cannot use a managed identity with an Azure Machine Learning compute cluster.

Azure Monitor

⚠ Warning

Azure Monitor supports using Azure Private Link to connect to a VNet. However, you must use the open Private Link mode in Azure Monitor. For more information, see Private Link access modes: Private only vs. Open.

Required public internet access

Azure Machine Learning requires both inbound and outbound access to the public internet. The following tables provide an overview of what access is required and what it is for. The **protocol** for all items is **TCP**. For service tags that end in `.region`, replace `region` with the Azure region that contains your workspace. For example, `Storage.westus`:

Direction	Ports	Service tag	Purpose
Inbound	29876-29877	BatchNodeManagement	Create, update, and delete of Azure Machine Learning compute instance and compute cluster. It isn't required if you use No Public IP option.
Inbound	44224	AzureMachineLearning	Create, update, and delete of Azure Machine Learning compute instance. It isn't required if you use No Public IP option.
Outbound	80, 443	AzureActiveDirectory	Authentication using Azure AD.

Direction	Ports	Service tag	Purpose
Outbound	443, 8787, 18881	AzureMachineLearning	Using Azure Machine Learning services.
Outbound	443	BatchNodeManagement.region	Communication with Azure Batch back-end for computes. Replace `region` with the Azure region of your workspace.
Outbound	443	AzureResourceManager	Creation of Azure resources with Azure Machine Learning.
Outbound	443, 445 (*)	Storage.region	Access data stored in the Azure Storage Account for compute cluster and compute instance. This outbound can be used to exfiltrate data. For more information, see Data exfiltration protection. (*) 445 is only required if you have a firewall between your virtual network for Azure ML and a private endpoint for your storage accounts.
Outbound	443	AzureFrontDoor.FrontEnd * Not needed in Azure China.	Global entry point for Azure Machine Learning studio. Store images and environments for AutoML.
Outbound	443	MicrosoftContainerRegistry.region **Note** that this tag has a dependency on the **AzureFrontDoor.FirstParty** tag	Access docker images provided by Microsoft. Setup of the Azure Machine Learning router for Azure Kubernetes Service.
Outbound	443	AzureMonitor	Used to log monitoring and metrics to App Insights and Azure Monitor.
Outbound	443	Keyvault.region	Access the key vault for the Azure Batch service. Only

Direction	Ports	Service tag	Purpose
			needed if your workspace was created with the hbi_workspace flag enabled.

Tip

If you need the IP addresses instead of service tags, use one of the following options:

- Download a list from Azure IP Ranges and Service Tags .
- Use the Azure CLI az network list-service-tags command.
- Use the Azure PowerShell Get-AzNetworkServiceTag command.

The IP addresses may change periodically.

Important

When using a compute cluster that is configured for **no public IP address**, you must allow the following traffic:

- **Inbound** from source of **VirtualNetwork** and any port source, to destination of **VirtualNetwork**, and destination port of **29876, 29877**.
- **Inbound** from source **AzureLoadBalancer** and any port source to destination **VirtualNetwork** and port **44224** destination.

You may also need to allow **outbound** traffic to Visual Studio Code and non-Microsoft sites for the installation of packages required by your machine learning project. The following table lists commonly used repositories for machine learning:

Host name	Purpose
anaconda.com *.anaconda.com	Used to install default packages.
*.anaconda.org	Used to get repo data.

Host name	Purpose
pypi.org	Used to list dependencies from the default index, if any, and the index isn't overwritten by user settings. If the index is overwritten, you must also allow ***.pythonhosted.org**.
cloud.r-project.org	Used when installing CRAN packages for R development.
***pytorch.org**	Used by some examples based on PyTorch.
***.tensorflow.org**	Used by some examples based on Tensorflow.
code.visualstudio.com	Required to download and install VS Code desktop. This is not required for VS Code Web.
update.code.visualstudio.com ***.vo.msecnd.net**	Used to retrieve VS Code server bits that are installed on the compute instance through a setup script.
marketplace.visualstudio.com **vscode.blob.core.windows.net** ***.gallerycdn.vsassets.io**	Required to download and install VS Code extensions. These enable the remote connection to Compute Instances provided by the Azure ML extension for VS Code, see Connect to an Azure Machine Learning compute instance in Visual Studio Code for more information.
raw.githubusercontent.com/microsoft/vscode-tools-for-ai/master/azureml_remote_websocket_server/*	Used to retrieve websocket server bits, which are installed on the compute instance. The websocket server is used to transmit requests from Visual Studio Code client (desktop application) to Visual Studio Code server running on the compute instance.

When using Azure Kubernetes Service (AKS) with Azure Machine Learning, allow the following traffic to the AKS VNet:

- General inbound/outbound requirements for AKS as described in the Restrict egress traffic in Azure Kubernetes Service article.
- **Outbound** to mcr.microsoft.com.
- When deploying a model to an AKS cluster, use the guidance in the Deploy ML models to Azure Kubernetes Service article.

For information on using a firewall solution, see Use a firewall with Azure Machine

Learning.

Secure the workspace with private endpoint

Azure Private Link lets you connect to your workspace using a private endpoint. The private endpoint is a set of private IP addresses within your virtual network. You can then limit access to your workspace to only occur over the private IP addresses. A private endpoint helps reduce the risk of data exfiltration.

For more information on configuring a private endpoint for your workspace, see How to configure a private endpoint.

> ⚠ **Warning**
>
> Securing a workspace with private endpoints does not ensure end-to-end security by itself. You must follow the steps in the rest of this article, and the VNet series, to secure individual components of your solution. For example, if you use a private endpoint for the workspace, but your Azure Storage Account is not behind the VNet, traffic between the workspace and storage does not use the VNet for security.

Secure Azure storage accounts

Azure Machine Learning supports storage accounts configured to use either a private endpoint or service endpoint.

Private endpoint

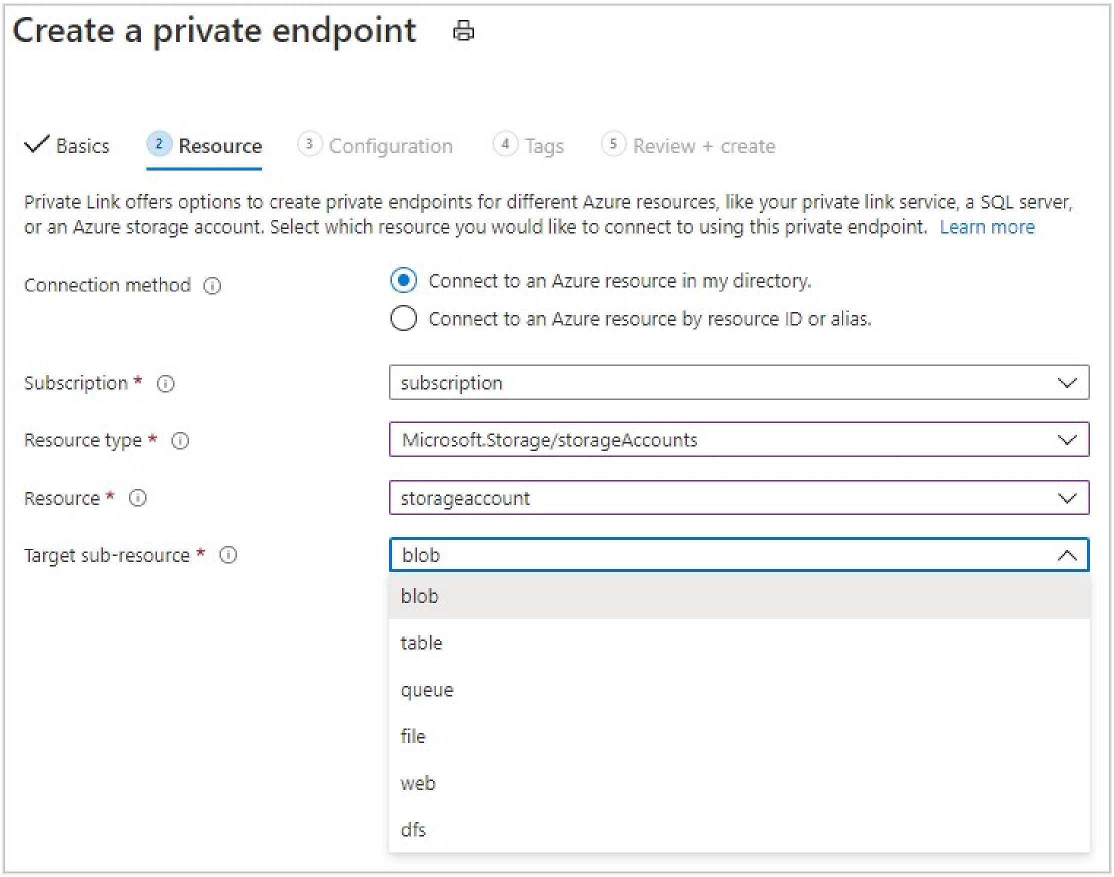
1. In the Azure portal, select the Azure Storage Account.

2. Use the information in Use private endpoints for Azure Storage to add private endpoints for the following storage resources:

   - **Blob**
   - **File**
   - **Queue** - Only needed if you plan to use ParallelRunStep in an Azure

Machine Learning pipeline.

- **Table** - Only needed if you plan to use ParallelRunStep in an Azure Machine Learning pipeline.



Create a private endpoint

Private Link offers options to create private endpoints for different Azure resources, like your private link service, a SQL server, or an Azure storage account. Select which resource you would like to connect to using this private endpoint. Learn more

Connection method ⓘ
● Connect to an Azure resource in my directory.
○ Connect to an Azure resource by resource ID or alias.

Subscription * ⓘ subscription

Resource type * ⓘ Microsoft.Storage/storageAccounts

Resource * ⓘ storageaccount

Target sub-resource * ⓘ blob

Tip

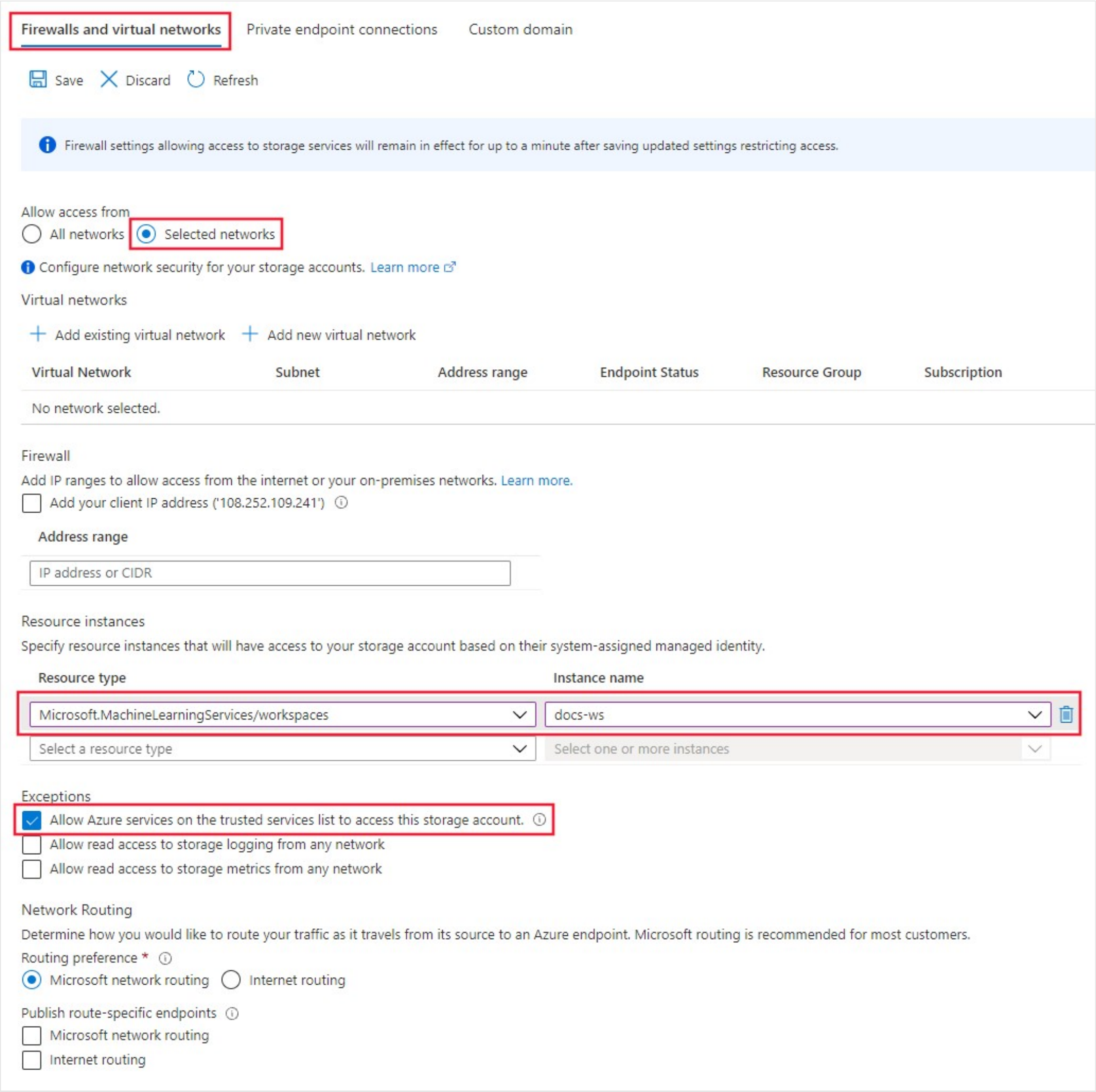
When configuring a storage account that is **not** the default storage, select the **Target subresource** type that corresponds to the storage account you want to add.

3. After creating the private endpoints for the storage resources, select the **Firewalls and virtual networks** tab under **Networking** for the storage account.

4. Select **Selected networks**, and then under **Resource instances**, select `Microsoft.MachineLearningServices/Workspace` as the **Resource type**. Select your workspace using **Instance name**. For more information, see Trusted access based on system-assigned managed identity.

Tip

Alternatively, you can select **Allow Azure services on the trusted services list to access this storage account** to more broadly allow access from trusted services. For more information, see Configure Azure Storage firewalls and virtual networks.



Firewall settings allowing access to storage services will remain in effect for up to a minute after saving updated settings restricting access.

Allow access from
○ All networks ● Selected networks

Configure network security for your storage accounts. Learn more

Virtual networks

+ Add existing virtual network + Add new virtual network

Virtual Network	Subnet	Address range	Endpoint Status	Resource Group	Subscription
No network selected.					

Firewall

Add IP ranges to allow access from the internet or your on-premises networks. Learn more.

□ Add your client IP address ('108.252.109.241')

Address range
IP address or CIDR

Resource instances

Specify resource instances that will have access to your storage account based on their system-assigned managed identity.

Resource type	Instance name
Microsoft.MachineLearningServices/workspaces	docs-ws
Select a resource type	Select one or more instances

Exceptions

☑ Allow Azure services on the trusted services list to access this storage account.

□ Allow read access to storage logging from any network

□ Allow read access to storage metrics from any network

Network Routing

Determine how you would like to route your traffic as it travels from its source to an Azure endpoint. Microsoft routing is recommended for most customers.

Routing preference *
● Microsoft network routing ○ Internet routing

Publish route-specific endpoints
□ Microsoft network routing
□ Internet routing

5. Select **Save** to save the configuration.

Tip

When using a private endpoint, you can also disable public access. For more information, see disallow public read access.

Secure Azure Key Vault

Azure Machine Learning uses an associated Key Vault instance to store the following credentials:

- The associated storage account connection string
- Passwords to Azure Container Repository instances
- Connection strings to data stores

Azure key vault can be configured to use either a private endpoint or service endpoint. To use Azure Machine Learning experimentation capabilities with Azure Key Vault behind a virtual network, use the following steps:

> **Tip**
>
> Regardless of whether you use a private endpoint or service endpoint, the key vault must be in the same network as the private endpoint of the workspace.

Private endpoint

For information on using a private endpoint with Azure Key Vault, see Integrate Key Vault with Azure Private Link.

Enable Azure Container Registry (ACR)

> **Tip**
>
> If you did not use an existing Azure Container Registry when creating the workspace, one may not exist. By default, the workspace will not create an ACR instance until it needs one. To force the creation of one, train or deploy a model using your workspace before using the steps in this section.

Azure Container Registry can be configured to use a private endpoint. Use the following steps to configure your workspace to use ACR when it is in the virtual network:

1. Find the name of the Azure Container Registry for your workspace, using one of

the following methods:

Azure CLI

APPLIES TO: ✓ Azure CLI ml extension **v2 (current)**

If you've installed the Machine Learning extension v2 for Azure CLI, you can use the `az ml workspace show` command to show the workspace information. The v1 extension does not return this information.

Azure CLI

```
az ml workspace show -w yourworkspacename -g resourcegroupname
--query 'container_registry'
```

This command returns a value similar to `"/subscriptions/{GUID}/resourceGroups/{resourcegroupname}/providers/Microsoft.ContainerRegistry/registries/{ACRname}"`. The last part of the string is the name of the Azure Container Registry for the workspace.

2. Limit access to your virtual network using the steps in Connect privately to an Azure Container Registry. When adding the virtual network, select the virtual network and subnet for your Azure Machine Learning resources.

3. Configure the ACR for the workspace to Allow access by trusted services.

4. Create an Azure Machine Learning compute cluster. This cluster is used to build Docker images when ACR is behind a VNet. For more information, see Create a compute cluster.

5. Use one of the following methods to configure the workspace to build Docker images using the compute cluster.

> ⓘ **Important**
>
> The following limitations apply When using a compute cluster for image builds:
>
> - Only a CPU SKU is supported.

- If you use a compute cluster configured for no public IP address, you must provide some way for the cluster to access the public internet. Internet access is required when accessing images stored on the Microsoft Container Registry, packages installed on Pypi, Conda, etc. You need to configure User Defined Routing (UDR) to reach to a public IP to access the internet. For example, you can use the public IP of your firewall, or you can use **Virtual Network NAT** with a public IP. For more information, see **How to securely train in a VNet**.

Azure CLI

You can use the `az ml workspace update` command to set a build compute. The command is the same for both the v1 and v2 Azure CLI extensions for machine learning. In the following command, replace `myworkspace` with your workspace name, `myresourcegroup` with the resource group that contains the workspace, and `mycomputecluster` with the compute cluster name:

Azure CLI

```
az ml workspace update --name myworkspace --resource-group myre-
sourcegroup --image-build-compute mycomputecluster
```

Tip

When ACR is behind a VNet, you can also **disable public access** to it.

Securely connect to your workspace

To connect to a workspace that's secured behind a VNet, use one of the following methods:

- Azure VPN gateway - Connects on-premises networks to the VNet over a private connection. Connection is made over the public internet. There are two types of VPN gateways that you might use:

  - Point-to-site: Each client computer uses a VPN client to connect to the VNet.
  - Site-to-site: A VPN device connects the VNet to your on-premises network.

- ExpressRoute - Connects on-premises networks into the cloud over a private connection. Connection is made using a connectivity provider.

- Azure Bastion - In this scenario, you create an Azure Virtual Machine (sometimes called a jump box) inside the VNet. You then connect to the VM using Azure Bastion. Bastion allows you to connect to the VM using either an RDP or SSH session from your local web browser. You then use the jump box as your development environment. Since it is inside the VNet, it can directly access the workspace. For an example of using a jump box, see Tutorial: Create a secure workspace.

> ⓘ **Important**
>
> When using a **VPN gateway** or **ExpressRoute**, you will need to plan how name resolution works between your on-premises resources and those in the VNet. For more information, see **Use a custom DNS server**.

If you have problems connecting to the workspace, see Troubleshoot secure workspace connectivity.

Workspace diagnostics

You can run diagnostics on your workspace from Azure Machine Learning studio or the Python SDK. After diagnostics run, a list of any detected problems is returned. This list includes links to possible solutions. For more information, see How to use workspace diagnostics.

Next steps

This article is part of a series on securing an Azure Machine Learning workflow. See the other articles in this series:

- Virtual network overview
- Secure the training environment

- Secure the inference environment
- Enable studio functionality
- Use custom DNS
- Use a firewall
- Tutorial: Create a secure workspace
- Tutorial: Create a secure workspace using a template
- API platform network isolation