Describe Azure ML Service best practices

Contents:

1. Manage and increase quotas for resources with Azure Machine Learning

Azure uses limits and quotas to prevent budget overruns due to fraud, and to honor Azure capacity constraints. Consider these limits as you scale for production workloads. This section will cover:

- Default limits on Azure resources related to Azure Machine Learning.

- Creating workspace-level quotas.

- Viewing your quotas and limits.

- Requesting quota increases.

- Private endpoint and DNS quotas.

Along with managing quotas, you can learn how to plan and manage costs for Azure Machine Learning.

Special Considerations:

- A quota is a credit limit, not a capacity guarantee. If you have large-scale capacity needs, contact Azure support to increase your quota.

- A quota is shared across all the services in your subscriptions, including Azure Machine Learning. Calculate usage across all services when you're evaluating capacity.

Azure Machine Learning compute is an exception. It has a separate quota from the core compute quota.

- Default limits vary by offer category type, such as free trial, pay-as-you-go, and virtual machine (VM) series (such as Dv2, F, and G).

Default resource quotas In this section, you learn about the default and maximum quota limits for the following resources:

- Virtual machines
- Azure Machine Learning compute
- Azure Machine Learning pipelines

- Azure Container Instances
- Azure Storage

Virtual Machines Each Azure subscription has a limit on the number of virtual machines across all services. Virtual machine cores have a regional total limit and a regional limit per size series. Both limits are separately enforced.

For example, consider a subscription with a US East total VM core limit of 30, an A series core limit of 30, and a D series core limit of 30. This subscription would be allowed to deploy 30 A1 VMs, or 30 D1 VMs, or a combination of the two that does not exceed a total of 30 cores.

You can't raise limits for virtual machines above the values shown in the following table.

Resource	Limit
Subscriptions per Azure Active Directory tenant	Unlimited
Coadministrators per subscription	Unlimited
Resource groups per subscription	980
Azure Resource Manager API request size	4,194,304 bytes
Tags per subscription1 50	
Unique tag calculations per subscription1	10,000
Subscription-level deployments per location	80022

---

1 You can apply up to 50 tags directly to a subscription. However, the subscription can contain an unlimited number of tags that are applied to resource groups and resources within the subscription. The number of tags per resource or resource group is limited to 50. Resource Manager returns a list of unique tag name and values in the subscription only when the number of tags is 10,000 or less. You still can find a resource by tag when the number exceeds 10,000.

2 If you reach the limit of 800 deployments, delete deployments that are no longer needed from the history. To delete subscription-level deployments, use Remove-AzDeployment or az deployment sub delete.

Azure Machine Learning Compute Azure Machine Learning compute has a default quota limit on both the number of cores and the number of unique compute resources allowed per region in a subscription. This quota is separate from the VM core quota from the previous section.

Request a quota increase to raise the limits in this section up to the maximum limit shown in the table.

Available resources:

- **Dedicated cores per region** have a default limit of 24 to 300, depending

on your subscription offer type. You can increase the number of dedicated cores per subscription for each VM family. Specialized VM families like NCv2, NCv3, or ND series start with a default of zero cores.

- **Low-priority cores per region** have a default limit of 100 to 3,000, depending on your subscription offer type. The number of low-priority cores per subscription can be increased and is a single value across VM families.

- **Clusters per region** have a default limit of 200. These are shared between a training cluster and a compute instance. (A compute instance is considered a single-node cluster for quota purposes.)

The following table shows additional limits that you can't exceed.

Resource	Maximum Limit
Workspaces per resource group	800
Nodes in a single Azure Machine Learning compute (AmlCompute) resource	100 nodes
GPU MPI processes per node	1-4
GPU workers per node	1-4
Job lifetime	21 days 1
Job lifetime on a low-priority node	7 days 2
Parameter servers per node	1

1 Maximum lifetime is the duration between when a run starts and when it finishes. Completed runs persist indefinitely. Data for runs not completed within the maximum lifetime is not accessible. 2 Jobs on a low-priority node can be preempted whenever there's a capacity constraint. We recommend that you implement checkpoints in your job.

Azure Machone Learning Pipelines   Azure Machine Learning Pipelines have the following limits.

Resource	Limit
Step in a pipeline	30,000
Workspaces per resource group	800

Container Instances   For more information, see Container Instances limits.

Storage   Azure Storage has a limit of 250 storage accounts per region, per subscription. This limit includes both Standard and Premium storage accounts.

To increase the limit, make a request through Azure Support.

Workspace-level quotas

Workspace-level quotas Use workspace-level quotas to manage Azure Machine Learning compute target allocation between multiple workspaces in the same subscription.

By default, all workspaces share the same quota as the subscription-level quota for VM families. However, you can set a maximum quota for individual VM families on workspaces in a subscription. This lets you share capacity and avoid resource contention issues.

1. Go to any workspace in your subscription.
2. In the left pane, select **Usages + quotas**.
3. Select the **Configure quotas** tab to view the quotas.
4. Expand a VM family.
5. Set a quota limit on any workspace listed under that VM family. You can't set a negative value or a value higher than the subscription-level quota.

Subscription *	Resource	Location *		
Azure ML Team Testing	Machine Learning Compute	West Europe		

Subscription View   Workspace View   Configure quotas

Resource name	Current limit	New limit		
> Standard D Family vCPUs	100			
> Standard DSv2 Family vCPUs	100			
∨ Standard Dv2 Family vCPUs	100			
azureml-dogbreeds (azureml-webinar)	10	10	💾	✅
azureml-webinar (azureml)	20	20	💾	✅
azuremlwbatchai_yopzkjia (azureml)	100	Unallocated cores: 0, Maximum: 100		
build19-weurope (build19)	100	Unallocated cores: 0, Maximum: 100		
danielsc (aml-workshop)	100	Unallocated cores: 0, Maximum: 100		
ignite2019 (ignite2019)	100	Unallocated cores: 0, Maximum: 100		
keras (azureml-webinar)	100	Unallocated cores: 0, Maximum: 100		
maxluk-azml (azureml-webinar)	100	Unallocated cores: 0, Maximum: 100		
pytorch-bert-squad (azureml-webinar)	100	Unallocated cores: 0, Maximum: 100		
test (amlworkspace)	100	Unallocated cores: 0, Maximum: 100		
test-sku-ws (azureml)	50	50	💾	✅
test2 (aml-workshop)	100	Unallocated cores: 0, Maximum: 100		
testws1 (azureml)	100	Unallocated cores: 0, Maximum: 100		
> Standard FSv2 Family vCPUs	100			

Note You need a subscription-level permissions to set a quote at the workspace level.

View your usage and quotas

To view your quota for various Azure resources like virtual machines, storage, or network, use the Azure portal:

1. On the left pane, select **All services** and then select **Subscriptions** under the **General** category.

2. From the list of subscriptions, select the subscription whose quota you're looking for.

3. Select **Usage + quotas** to view your current quota limits and usage. Use the filters to select the provider and locations.

You manage the Azure Machine Learning compute quota on your subscription separately from other Azure quotas:

1. Go to your **Azure Machine Learning** workspace in the Azure portal.

2. On the left pane, in the **Support + troubleshooting** section, select **Usage + quotas** to view your current quota limits and usage.

3. Select a subscription to view the quota limits. Filter to the region you're interested in.

4. You can switch between a subscription-level view and a workspace-level view.

Request quota increases   To raise the limit or quota above the default limit, open an online customer support request at no charge.

You can't raise limits above the maximum values shown in the preceding tables. If there's no maximum limit, you can't adjust the limit for the resource.

When you're requesting a quota increase, select the service that you have in mind. For example, select Azure Machine Learning, Container Instances, or Storage. For Azure Machine Learning compute, you can select the **Request Quota** button while viewing the quota in the preceding steps.

Private endpoint and private DNS quota increases   There are limits on the number of private endpoints and private DNS zones that you can create in a subscription.

Azure Machine Learning creates resources in your (customer) subscription, but some scenarios create resources in a Microsoft-owned subscription.

In the following scenarios, you might need to request a quota allowance in the Microsoft-owned subscription:

- Azure Private Link enabled workspace with a customer-managed key (CMK)
- Azure Container Registry for the workspace behind your virtual network
- Attaching a Private Link enabled Azure Kubernetes Service cluster to your workspace

To request an allowance for these scenarios, use the following steps:

1. Create an Azure support request and select the following options in the **Basics** section:

Field	Selection
Issue type	**Techical**
Service	**My services**. Then select **Machine Learning** in the drop-down list.
Problem type	**Workspace Configuration and Security**
Problem subtype	**Private Endpoint and Private DNS Zone allowance request**

2. In the **Details** section, use the **Description** field to provide the Azure region and the scenario that you plan to use. If you need to request quote increases for multiple subscriptions list the sunscriptions IDs in this field.

3. Select *Create* to create the request.

Additional considerations

- Create different workspaces by different department / business team / data tier, and per environment (development, staging, and production) - across

relevant Azure subscriptions

- Define workspace level tags which propagate to initially provisioned resources in managed resource group (Tags could also propagate from parent resource group)
- Use Azure Resource Manager templates templates to have a more managed way of deploying the workspaces - whether via CLI, PowerShell, or some SDK
- Create relevant groups of users - using Group REST API or by using Azure Active Directory Group Sync with SCIM.

2. Understand workspace administration best practices

Azure subscription limits

Key Azure limits are:

- Storage accounts per region per subscription: **250**
- Maximum egress for general-purpose v2 and Blob storage accounts (all regions): **50 Gbps**
- Virtual Machines (VMs) per subscription per region: **25,000**
- Resource groups per subscription: **980**
- These limits are at this point in time and might change going forward. Some of them can also be increased if needed. For more help in understanding the impact of these limits or options of increasing them, please contact Microsoft or Databricks technical architects.

Networking

When creating a workspace the default network configuration is to use a **Public endpoint**, which is accessible on the public internet. To limit access to your workspace to an Azure Virtual Network you have created, you can instead select **Private endpoint** (preview) as the **Connectivity method**, and then use **+ Add** to configure the endpoint.

Machine Learning
Create a machine learning workspace

Basics   **Networking**   Advanced   Tags   Review + create

Network connectivity

You can connect to your workspace either publicly or privately using a private endpoint.

Connectivity method *   ○ Public endpoint (all networks)
     ● Private endpoint

Private endpoint

Create a private endpoint to allow a private connection to this resource.

Name	Subscription	Resource group	Region	Subnet	Private DNS Zone
Click on add to create a private endpoint					

+ Add

Review + create   < Previous   Next : Advanced

On the **Create private endpoint** form, set the location, name, and virtual network to use. If you'd like to use the endpoint with a Private DNS Zone, select **Integrate with private DNS** zone and select the zone using the **Private DNS Zone field**. Select **OK** to create the endpoint.

8

Create private endpoint   ×

Subscription * ⓘ   documentationteam   ⌄

  Resource group * ⓘ   myresourcegroup   ⌄
                          Create new

Location *   (US) Central US   ⌄

Name * ⓘ   myendpoint   ✓

Workspace sub-resource * ⓘ amlworkspace   ⌄

Networking

To deploy the private endpoint, select a virtual network subnet. Learn more about private endpoint networking ⬈

Virtual network * ⓘ   myvnet   ⌄

Subnet * ⓘ   default (172.17.0.0/24)   ⌄

                          ⓘ If you have a network security group (NSG) enabled for the subnet above, it will be disabled
                          for private endpoints on this subnet only. Other resources on the subnet will still have NSG
                          enforcement.

Private DNS integration

To connect privately with your private endpoint, you need a DNS record. We recommend that you integrate your private endpoint with a
private DNS zone. You can also utilize your own DNS servers or create DNS records using the host files on your virtual machines.
Learn more about private DNS integration ⬈

Integrate with private DNS zone ⓘ   [**Yes**   No]

Private DNS Zone * ⓘ   (New) privatelink.workspacecore.azure.net   ⌄

[**OK**] [Discard]

When you are finished configuring networking, you can select **Review + Create**,
or advance to the optional **Advanced** configuration. **Important**: Using a
private endpoint with Azure Machine Learning workspace is currently in public
preview. This preview is provided without a service level agreement, and it's
not recommended for production workloads. Certain features might not be
supported or might have constrained capabilities. For more information, see
Supplemental Terms of Use for Microsoft Azure Previews.

Multiple workspaces with private endpoint

When you create a private endpoint, a new Private DNS Zone named **pri-
vatelink.api.azureml.ms** is created. This contains a link to the virtual net-
work. If you create multiple workspaces with private endpoints in the same
resource group, only the virtual network for the first private endpoint may be
added to the DNS zone. To add entries for the virtual networks used by the
additional workspaces/private endpoints, use the following steps:

1. In the Azure portal, select the resource group that contains the
   workspace. Then select the Private DNS Zone resource named
   **privatelink.api.azureml.ms**

2. In the **Settings**, select **Virtual network links**.
3. Select **Add**. From the **Add virtual network link** page, provide a unique **Link name**, and then select the **Virtual network** to be added. Select **OK** to add the network link. For more information, see Azure Private Endpoint DNS configuration.

Vulnerability scanning

Azure Security Center provides unified security management and advanced threat protection across hybrid cloud workloads. You should allow Azure Security Center to scan your resources and follow its recommendations. For more, see Azure Container Registry image scanning by Security Center and Azure Kubernetes Services integration with Security Center.

Download a configuration file

If you will be creating a compute instance, skip this step. The compute instance has already created a copy of this file for you.
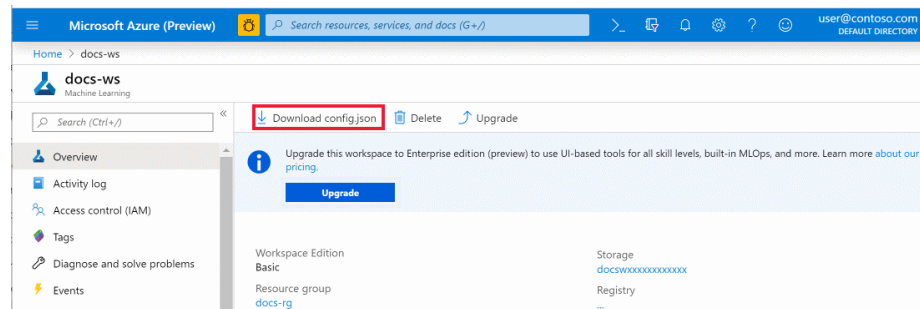
Place the file into the directory structure with your Python scripts or Jupyter Notebooks. It can be in the same directory, a subdirectory named *.azureml*, or in a parent directory. When you create a compute instance, this file is added to the correct directory on the VM for you.

3. Security best practices

Security and infrastructure configuration go hand-in-hand. When you set up your Azure ML service workspace(s) and related services, you need to make sure that security considerations do not take a back seat during the architecture design.

Always hide secrets in a key vault

It is a significant security risk to expose sensitive data such as access credentials openly in Notebooks or other places such as job configs, initialization scripts, etc. You should always use a vault to securely store and access them. Unless you

are using azure Databricks (ADB) workspace in which case you can use ADB's internal Key Vault for this purpose, use Azure's Key Vault (AKV) service.

If using Azure Key Vault, create separate AKV-backed secret scopes and corresponding AKVs to store credentials pertaining to different data stores. This will help prevent users from accessing credentials that they might not have access to. Since access controls are applicable to the entire secret scope, users with access to the scope will see all secrets for the AKV associated with that scope.

More Information can be found in the following urls:

Create an Azure Key Vault-backed secret scope

Example of using a secret in a notebook

Best practices for creating secret scopes

Additional considerations

Configure encryption-at-rest for Blob Storage and ADLS, preferably by using customer-managed keys in Azure Key Vault.

4. Tools and integration best practices

- Use Azure Data Factory to orchestrate pipelines / workflows (or something like Airflow).
- Sync notebooks with Azure DevOps for seamless version control. For detailed information refer to his page.