

# Compute Services on the Cloud

## Working With Images

- Azure VM Image is snapshot of managed disk + additional metadata
- In Azure VM images we have two types of VM images
  - Generalized VM images : These images will allow the user to set username and passwords/keys during creation
  - Specialized VM images : same as AMI (usernames are fixed)
- Azure while creating the image will remove all the user specific information from vm and then taken a snapshot
- So once we use a vm to create a generalized vm image, vm becomes unusable.
- Azure VM images also are of two types
  - free
  - paid
- Azure also has market place
- Azure VM image has four properties
  - publisher : Organization/individual who is publishing image
  - offer : product name/ server name
  - sku : version of the product/ different releases
  - Version: current versions
- Azure vm image has a unique id or URN :::
- To store vm images we create a compute gallery with a vm image definition and a specific image version.
- image definitions are like to store an application with different versions

### Activity

- create a new resource group for this activity
- Create a new virtual Machine and install nginx/apache and host a website by following below instructions
- Use ubuntu machine

```
#!/bin/bash
sudo apt update
sudo apt install nginx unzip -y
cd /tmp && wget https://www.free-css.com/assets/files/free-css-templates/download/page294/troweld.zip && unzip troweld.zip
```

```
sudo mv /tmp/troweld-html /var/www/html/troweld
```

## Create a virtual machine ...


[Help me create a low cost VM](#)
[Help me create a VM optimized for high availability](#)
[Help me choose the right VM size for my workload](#)

[Basics](#)
[Disks](#)
[Networking](#)
[Management](#)
[Monitoring](#)
[Advanced](#)
[Tags](#)
[Review + create](#)

Add additional configuration, agents, scripts or applications via virtual machine extensions or cloud-init.

### Extensions

Extensions provide post-deployment configuration and automation.

Extensions ⓘ

[Select an extension to install](#)

### VM applications

VM applications contain application files that are securely and reliably downloaded on your VM after deployment. In addition to the application files, an install and uninstall script are included in the application. You can easily add or remove applications on your VM after create. [Learn more](#) ⓘ


[Select a VM application to install](#)

### Custom data and cloud init

Pass a cloud-init script, configuration file, or other data into the virtual machine **while it is being provisioned**. The data will be saved on the VM in a known location. [Learn more about custom data for VMs](#) ⓘ

Custom data

```
#!bin/bash
sudo apt update
sudo apt install nginx unzip -y
cd /tmp && wget https://www.free-css.com/assets/files/free-css-templates/download/page294/troweld.zip && unzip troweld.zip
sudo mv /tmp/troweld-html /var/www/html/troweld
```

 Custom data on the selected image will be processed by cloud-init.
 [Learn more about custom data for VMs](#) ⓘ

### User data

Pass a script, configuration file, or other data that will be accessible to your applications throughout the lifetime of the virtual machine. Don't use user data for storing your secrets or passwords. [Learn more about user data for VMs](#) ⓘ

Enable user data

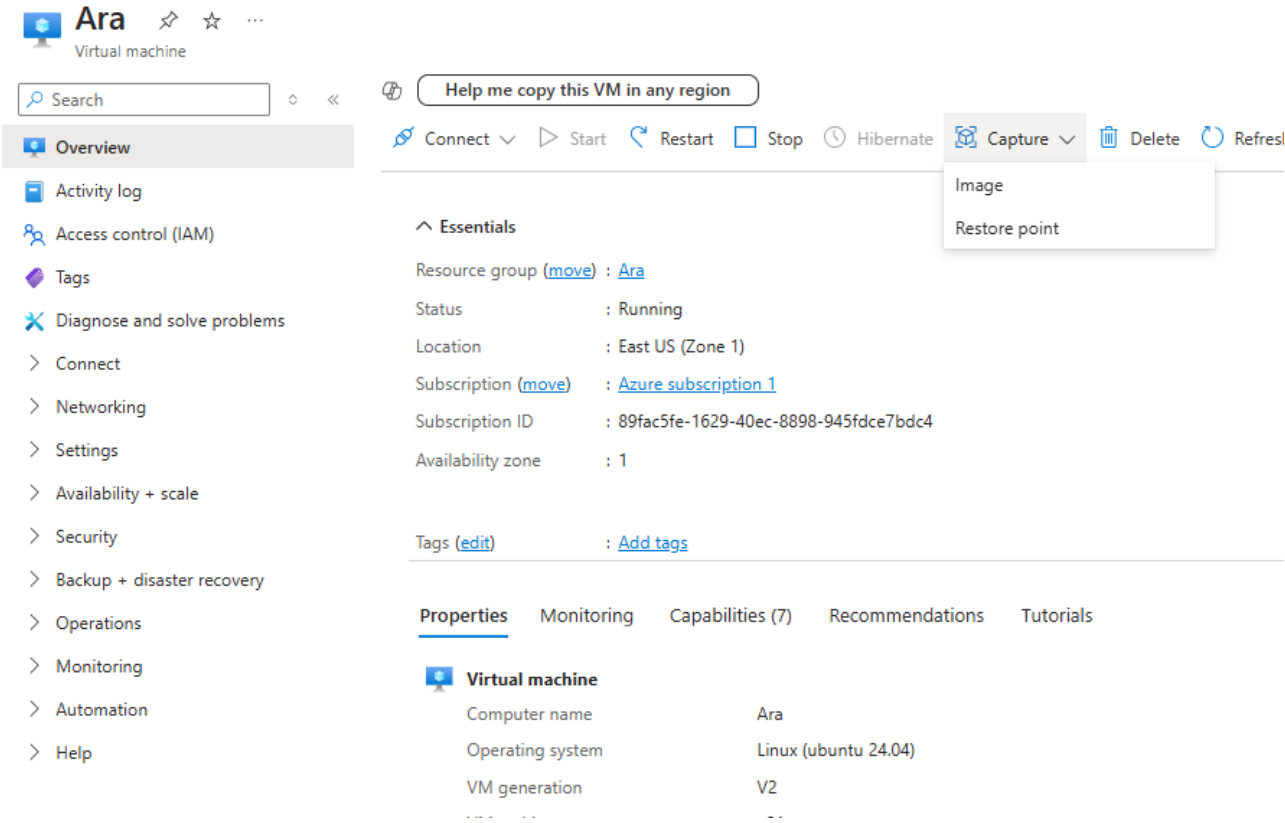
☐

- Write above script in Advanced tab while creating virtual machine and create VM
- Now verify the website and connect to VM once to check connection

### Taking Image

- Go to Virtual Machines and select the created Virtual Machine

- in the horizontal blade find **Capture** select Capture and click on Image



Create an image ...

Cancel

4 / 36

Home > CreateVm-canonical.ubuntu-24\_04-lts-server-20241119174029 | Overview > Ara >

## Create an image

Basics Tags Review + create

Create an image from this virtual machine that can be used to deploy additional virtual machines and virtual machine scale sets. With a shared image, you can easily replicate the image to Azure regions around the world and manage versions of the image. Certain information from the virtual machine will be carried forward to the image including OS type, VM generation, plan, and publishing details. [Learn more](#)

### Project details

Subscription Azure subscription 1

Resource group \* Ara

### Instance details

Region (US) East US

Share image to Azure compute gallery ☒ Yes, share it to a gallery as a VM image version.  
☐ No, capture only a managed image.

**i** Managed image is not available because it is not currently supported with Trusted launch virtual machines.

Automatically delete this virtual machine after creating the image ☐

### Gallery details

Target Azure compute gallery \* (new) Websites

[Create new](#)

Operating system state ☒ Generalized: VMs created from this image require hostname, admin user, and other VM related setup to be completed on first boot  
☐ Specialized: VMs created from this image are completely configured and do not require parameters such as hostname and admin user/password

- Create a New Gallery and select Generalized
- Generalized : if you select this the user data from the VM is deleted and takes the snapshot. provides option to create a new Vm with new user details
- Specialized : if you select this , all the data will remain same including user details

Microsoft Azure

Home > CreateVm-canonical.ubuntu-24\_04-lts-server-20241119174029 | Overview > Ara >

### Create an image

Operating system state ☒ Generalized: VMs created from this image require hostname, admin user, and other VM related setup to be completed on first boot  
☐ Specialized: VMs created from this image are completely configured and do not require parameters such as hostname and admin user/password

**Warning:** Capturing a virtual machine image will make the virtual machine unusable. This action cannot be undone.

Target VM image definition \* Create a VM image definition

[Create new](#)

**Version details**

Version number \* Example: 0.0.1, 15.35.0

Exclude from latest ☐

End of life date DD/MM/YYYY

Lock deleting Replicated Locations ☒

Shallow replication ☐

**Replication**

A VM image version can be replicated to different regions depending on what makes sense for your organization. One example is to always replicate the latest image in multiple regions while all older versions are only available in 1 region. This can help save on storage costs for VM image versions.

Default storage sku Zone-redundant

Default replica count \* 1

Target regions	Replica count	Storage sku
(US) East US	1	Zone-redundant
	1	Standard HDD LRS

### Create a VM image definition

VM image definition name \*

OS type ☒ Linux  
☐ Windows

VM generation ☐ Gen 1  
☒ Gen 2

Security type Trusted launch

VM architecture ☒ x64  
☐ Arm64

**i** VM architecture has been automatically switched to x64 because Arm64 virtual machines are not supported with Trusted and Confidential security type.

Higher storage performance with NVMe ☐

Hibernation supported ☐

Accelerated networking ☐

Publisher \* canonical

Offer \* ubuntu-24\_04-lts

SKU \* server

**Publishing options (Optional)**

- Select Create new **image definition** fill all the details

Home > CreateVM-canonical.ubuntu-24.04-ls-server-20241119174029 | Overview > Ara >

Create an image ...

Operating system state ⓘ

☒ Generalized: VMs created from this image require hostname, admin user, and other VM related setup to be completed on first boot

☐ Specialized: VMs created from this image are completely configured and do not require parameters such as hostname and admin user/password

⚠ Capturing a virtual machine image will make the virtual machine unusable. This action cannot be undone.

Target VM image definition ⓘ

Create a VM image definition

Create new

Version details

Version number ⓘ

1.0.0

Exclude from latest ⓘ

☐

End of life date ⓘ

DD/MM/YYYY

Lock deleting Replicated Locations ⓘ

☒

Shallow replication ⓘ

☐

Replication

A VM image version can be replicated to different regions depending on what makes sense for your organization. One example is to always replicate the latest image in multiple regions while all older versions are only available in 1 region. This can help save on storage costs for VM image versions.

Default storage sku ⓘ

Zone-redundant

Default replica count ⓘ

1

Target regions

Replica count

Storage sku

(US) East US

1

Zone-redundant

1

Standard HDD LRS

Review + create

< Previous

Next : Tags >

Create a VM image definition

VM image definition name ⓘ

Website

OS type ⓘ

☒ Linux

☐ Windows

VM generation ⓘ

☐ Gen 1

☒ Gen 2

Security type ⓘ

Trusted launch

VM architecture ⓘ

☒ x64

☐ Arm64

VM architecture has been automatically switched to x64 because Arm64 virtual machines are not supported with Trusted and Confidential security type.

Higher storage performance with NVMe ⓘ

☐

Hibernation supported ⓘ

☐

Accelerated networking ⓘ

☐

Publisher ⓘ

Aravindh

Offer ⓘ

ubuntu\_website

SKU ⓘ

server

Publishing options (Optional)

Ok

Cancel

Give feedback

# Create an image ...

✓ Validation passed

BasicsTagsReview + create

Basics

Subscription	Azure subscription 1
Resource group	Ara
Region	East US
Share image to Azure compute gallery	Yes
Automatically delete this virtual machine after creating the image	No
Azure compute gallery	(new) Websites
Operating system state	Generalized
Target VM image definition	(new) Website
Version number	1.0.0
Source virtual machine	Ara
Exclude from latest	No
End of life date	None
Lock deleting Replicated Locations	Yes
Shallow replication	No

Replication

Default replica count	1
Replication	East US: 1

Tags

(none)

Create

< Previous

Next >

Download a template for automation

- Create image
- This will take a decent time to crete VM , as we selected Generalized image
- the VM won't work delete the VM check the website and connection to vm for confirmation
- After deletion of VM, Create a new VM

- while selecting a new VM image select on [see all images](#)

**Availability zone** \* ⓘ Zone 1

Using an Azure-selected zone is not supported in region 'East US'.

You can now select multiple zones. Selecting multiple zones will create one VM per zone. [Learn more](#)

**Security type** ⓘ Trusted launch virtual machines

[Configure security features](#)

**Image** \* ⓘ Ubuntu Server 24.04 LTS - x64 Gen2

[See all images](#) [Configure VM generation](#)

**VM architecture** ⓘ ☐ Arm64 ☒ x64

**Run with Azure Spot discount** ⓘ ☐

**Size** \* ⓘ Standard\_B1s - 1 vcpu, 1 GiB memory (₹631.61/month) (free services eligible)

[See all sizes](#)

**Enable Hibernation** ⓘ ☐

Hibernation does not currently support Trusted launch and Confidential virtual machines for Linux images. [Learn more](#)

#### Administrator account

[Home](#) > [Virtual machines](#) > [Create a virtual machine](#) >

### Select an image

#### Other Items

##### My Images

##### Shared Images

##### Community Images

##### Direct Shared Images (PREVIEW)

#### Marketplace

##### All

##### Recently created

##### Private products

#### Categories

Compute (3774)

### Other Items | Shared Images

Search in Shared Images

Publisher : **All**

Azure Compute Gallery : **All**

Image Name	↑↓	Subscription	↑↓	Publisher
No results				

- here select shared you will find the image you created above select it and create VM.
- After creation of VM, take public ip and check whether website is working or not and check SSH connection from powershell

## Working with Scaling

- Scaling refers to increasing machine sizes or machines
- Scaling is of two types
  - Vertical Scaling
  - Horizontal Scaling
- Vertical Scaling : This refers to increase in size of CPU/RAM or even instance families in cloud

## Activity



- Open **Virtual Machine Scale Sets** and click on Create New

Microsoft Azure

Search resources, services, and docs (G+/I)

Copilot

[Home](#) > [Virtual machine scale sets](#) >

## Create a virtual machine scale set ...

Basics

Spot

Disks

Networking

Management

Health

Advanced

Tags

Review + create

Azure virtual machine scale sets let you create and manage a group of load balanced VMs. The number of VM instances can automatically increase or decrease in response to demand or a defined schedule. Scale sets provide high availability to your applications, and allow you to centrally manage, configure, and update a large number of VMs.  
[Learn more about virtual machine scale sets](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*

Azure subscription 1

Resource group \*

vmssdemo

[Create new](#)

### Scale set details

Virtual machine scale set name \*

Trowelds

Region \*

(US) East US

Availability zone ⓘ

Zones 1, 2

Autoscaling can help you respond to an outage by scaling out new instances in another zone.

### Orchestration

A scale set has a "scale set model" that defines the attributes of virtual machine instances (size, number of data disks, etc). As the number of instances in the scale set changes, new instances are added based on the scale set model.  
[Learn more about the scale set model](#)

Orchestration mode \* ⓘ

Flexible: achieve high availability at scale with identical or multiple virtual machine types

Uniform: optimized for large scale stateless workloads

Security type ⓘ

Trusted launch virtual machines

[Configure security features](#)

< Previous

Next : Spot >

Review + create

- Provide Resource group name
- Give Scale Sets Name
- Region make sure your image is there in same region
- select zones
- Orchestration:
  - Flexible : you can add different types of Vm's with different sizes
  - Uniform : Here it is Same Virtual Machines with sizes


[Home](#) > [Virtual machine scale sets](#) >

## Create a virtual machine scale set ...

### Orchestration


A scale set has a "scale set model" that defines the attributes of virtual machine instances (size, number of data disks, etc). As the number of instances in the scale set changes, new instances are added based on the scale set model.

[Learn more about the scale set model](#) 

Orchestration mode \* 

☐ **Flexible:** achieve high availability at scale with identical or multiple virtual machine types


☒ **Uniform:** optimized for large scale stateless workloads

Security type 

Trusted launch virtual machines 

[Configure security features](#)


### Scaling

Scaling mode 

☒ **Manually update the capacity:** Maintain a fixed amount of instances.

☐ **Autoscaling:** Scaling based on a CPU metric, on any schedule.


☐ **No scaling profile:** manual attach virtual machines after deployment



Instance count \* 

2 

[Configure scaling options](#)

### Instance details

Image \* 

 Ubuntu Server 24.04 LTS - x64 Gen2 


[See all images](#) | [Configure VM generation](#)

VM architecture 

☐ Arm64

☒ x64



Size \* 

Standard\_B1s - 1 vcpu, 1 GiB memory (₹631.61/month) (free services eligible) 


[See all sizes](#)

Enable Hibernation 

☐

 Hibernate does not currently support Uniform Orchestration mode. [Learn more](#) 

### Administrator account

Authentication type 

☐ Password

[< Previous](#)

[Next : Spot >](#)

[Review + create](#)

- Scaling :
  - Manually update the Capacity : Maintain a fixed amount of instances
  - Autoscaling : Scaling based on a CPU metric, on any schedule.
- **Select Manually update the capacity option**

- Select instance count = 2

## Create a virtual machine scale set ...

Scaling mode ⓘ

☒ Manually update the capacity: Maintain a fixed amount of instances.

☐ Autoscaling: Scaling based on a CPU metric, on any schedule.

☐ No scaling profile: manual attach virtual machines after deployment

Instance count \* ⓘ

2 ✓

[Configure scaling options](#)

**Instance details**

Image \* ⓘ

Ubuntu Server 24.04 LTS - x64 Gen2 ✓

[See all images](#) | [Configure VM generation](#)

VM architecture ⓘ

☐ Arm64

☒ x64

Size \* ⓘ

Standard\_B1s - 1 vcpu, 1 GiB memory (₹631.61/month) (free services eligible) ✓

[See all sizes](#)

Enable Hibernation ⓘ

☐

**i** Hibernate does not currently support Uniform Orchestration mode. [Learn more](#) ↗

**Administrator account**

Authentication type ⓘ

☐ Password

☒ SSH public key

Username \* ⓘ

Ara ✓

SSH public key source

Use existing key stored in Azure ✓

**i** Ed25519 and RSA SSH formats are supported for the selected VM image. Ed25519 provides a fixed security level of no more than 128 bits for 256-bit key, while RSA could offer better security with keys longer than 3072 bits.

☒

Stored Keys \*

syskey ✓

- image : click on see all images
  - under **shared** select your desired image
- size :
  - select your desired size
- username : someusername
- Authentication type : choose one of them Password/SSH public key
- License type : Other
- click on Next (Spot)
- **SPOT VM's** Spot VM's are defined as, where there is less demand for VM's Azure offers higher size VM's for low cost, these are not permanent. This will be taken back without any intimation at any time. There is risk involved.

- This section provides to utilize SPOT VM's which comes for low cost
- These VM's will be taken back at any time.
- Skip this and go to next section

Disks :

- Change OS Disk type = standard SSD to decrease our cost

Networking :

- Select the virtual network if you have created, make sure the image should be there in same region
- select the **edit** icon for NIC card
  - Enable public ip [this is to verify in our activity]
  - select inbound ports
    - SSH
    - HTTP
- Select **None** for LoadBalancing options

Management :

Management tells about upgrading of VM's

The Upgrade Policy in Azure Virtual Machine Scale Sets (VMSS) allows you to control how updates to your instances are applied. You can either manage upgrades manually, let Azure apply them automatically, or use a rolling upgrade strategy to apply changes incrementally across your instances. The rolling upgrade option offers the best flexibility for critical applications, allowing for high availability during the upgrade process.

- the options speaks about answer for below question
  - how do you want to upgrade your machines when a new version of image comes in
  - Automatic : (Azure will take over this)
    - when you give a new image Azure will automatically update all the machines with new image without any downtime.
  - Manual :
    - you have to click on some button to happen this upgrade
  - Rolling :

- this option is for upgrading batch wise

#### Microsoft Defender for Cloud

Microsoft Defender for Cloud provides unified security management and advanced threat protection across hybrid cloud workloads. [Learn more](#)

✓ Your subscription is protected by Foundational Cloud Security Posture Management Free Plan.

#### Upgrade policy

Upgrade mode \* ⓘ Rolling - Upgrades roll out in batches with optional pause

Rolling upgrade batch size % \* 20

Pause time between batches (sec) \* 2

Max unhealthy instance % \* 20

Max unhealthy upgraded % \* 20

Prioritize unhealthy instances ⓘ ☐

Enable cross-zone upgrade ⓘ ☐

- Select **Automatic** and go to next section

Health :

- Enable Application health monitoring
  - this will check your application is active or not
  - Configure :
    - you can give a path here, so that you can check whether the page is responsive or not.
  - if the page is not responsive for the port 80 , using Automatic repair enabled, the vm will be deleted and creates a new

Basics Spot Disks Networking Management **Health** Advanced Tags Review + create


You can configure health monitoring on an application endpoint to update the status of the application on that instance. This instance status is required to enable platform managed upgrades like automatic OS updates and virtual machine instance upgrades. [Learn more about application health monitoring](#)

### Health

Enable application health monitoring ☒

Health monitor configuration

Type: Application health extension  
Protocol: HTTP  
Port number: 80  
Path: /  
[Configure](#)

 The Application Health extension will probe the application health endpoint and update the status of the application. When the health endpoint is not set up correctly the status of the application will be reported as unhealthy. [Learn more](#)

### Recovery

Azure can help you recover your instances from control plane and application failures. Before enabling the automatic repairs policy, review the requirements for opting in. [Learn more about automatic repair policy](#)

Automatic repairs ☒

Repair actions

Replace

Grace period (min) \*

10

Advanced :

- How you want your scaling to be done

Basics Spot Disks Networking Management Health **Advanced** Tags Review + create

Add additional configuration, agents, scripts or applications via virtual machine extensions or cloud-init.

### Allocation policy

Enable scaling beyond 100 instances ☒

Force strictly even balance across zones ☒


Spreading algorithm

☒ Max spreading  
☐ Fixed spreading

- select **Force strictly even balance across zones** :
  - this option make sure that if you hve 10 machines it will maintain 5 in one zone and 5 in another zone.
- Spreading algorithm:
  - Max spreading :

- fixed spreading
  - for fixed you need to create `fault domain` which is not recommended.
- Review and create it.
- Once deployment is succesfull
- Open the virtul machine scalesets

[Home](#) > [Virtual machine scale sets](#) > [abcd-troweld-dev-vmss](#)

**abcd-troweld-dev-vmss** | Instances

Virtual machine scale set

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

**Instances**

Networking

Settings

Availability + scale

- Scaling
- Availability
- Size

Security

**Operations**

Monitoring

- Insights
- Alerts
- Metrics
- Logs
- Connection monitor
- Workbooks

Automation


Help

Start

Restart

Stop

	Instance	↑↓	Computer i
<input type="checkbox"/>	abcd-troweld-dev-v...		abcd-trow0
<input type="checkbox"/>	abcd-troweld-dev-v...		abcd-trow0
<input type="checkbox"/>	abcd-troweld-dev-v...		abcd-trow0

 When an instance returns 200 (OK) on Healthy. In all other cases it shows Unl

- from the left side balde select **Instances**
- check the number of instances the count will be **2**
- As we selected **Manual update capacity** and the count we gave 2, since it created 2 machines.
- login to both machines

- use public ip and login from powershell
- login : `ssh Ara@ip`
- after logged in
- install Stress applplication, which increses stress on cpu and shows 100%.
- cmd : `sudo apt install stress`
- after installation is finished , type cmd : `stress`

```
Ara@abcd-trow000000:~$ stress
'stress' imposes certain types of compute stress on your system

Usage: stress [OPTION [ARG]] ...
  -?, --help            show this help statement
  --version             show version statement
  -v, --verbose         be verbose
  -q, --quiet           be quiet
  -n, --dry-run         show what would have been done
  -t, --timeout N       timeout after N seconds
  --backoff N           wait factor of N microseconds before work starts
  -c, --cpu N           spawn N workers spinning on sqrt()
  -i, --io N            spawn N workers spinning on sync()
  -m, --vm N            spawn N workers spinning on malloc()/free()
  --vm-bytes B          malloc B bytes per vm worker (default is 256MB)
  --vm-stride B         touch a byte every B bytes (default is 4096)
  --vm-hang N           sleep N secs before free (default none, 0 is inf)
  --vm-keep             redirty memory instead of freeing and reallocating
  -d, --hdd N           spawn N workers spinning on write()/unlink()
  --hdd-bytes B         write B bytes per hdd worker (default is 1GB)

Example: stress --cpu 8 --io 4 --vm 2 --vm-bytes 128M --timeout 10s

Note: Numbers may be suffixed with s,m,h,d,y (time) or B,K,M,G (size).
Ara@abcd-trow000000:~$ |
```

- copy the hilighed line : `stress --cpu 8 --io 4 --vm 2 --vm-bytes 128M --timeout 15m -v`
- change the timing at last
- do this on both the machines.
- Above activity increses load on both machines
- As we enabled Manual the VM'd wont ge incresed based on any factor the count remains same. *Even if one of the two VMs is deleted, a new VM will not be created to maintain the desired count. This feature is not available in Azure but is supported in AWS.*
- To enable this feature, you need to change the update policy to **Automatic** and set the
- Now go to **Availability+ scale** section



- select Scaling option

Home > Virtual machine scale sets > abcd-troweld-dev-vmss

**abcd-troweld-dev-vmss | Scaling**

Virtual machine scale set

Search Save Discard Refresh Logs Feedback

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Instances Networking Settings **Availability + scale** Scaling Availability Size Security Operations Monitoring Insights Alerts Metrics Logs Connection monitor Workbooks Automation Help

**Configure** Scale-In Policy Predictive charts Run history JSON Notify Diagnostic settings

Autoscale is a built-in feature that helps applications perform their best when demand changes. You can choose to scale your resource manually to a specific instance count, or via a custom Autoscale policy that scales based on metric(s) thresholds, or schedule instance count which scales during designated time windows. Autoscale enables your resource to be performant and cost effective by adding and removing instances based on demand. [Learn more about Azure Autoscale](#) or [view the how-to video](#).

**Choose how to scale your resource**

**Manual scale** ☐ Maintain a fixed instance count

**Custom autoscale** ☒ Scale on any schedule, based on any metrics or predictively

**Custom autoscale**

Autoscale setting name abcd-troweld-dev-vmss-Autoscale-297

Resource group abcd-troweld-dev-vmss\_group

Instance count ① 3

Predictive autoscale Mode Disabled ▾ Pre-launch setup of instances (minutes) ①

Enable Forecast only or Predictive autoscale. [Learn more about Predictive autoscale.](#)

**Default\*** Auto created default scale condition ✎

Delete warning *The very last or default recurrence rule cannot be deleted. Instead, you can disable autoscale to turn off autoscale.*


Scale mode ☒ Scale based on a metric ☐ Scale to a specific instance count

- right side you can see two options


- Manual Scale
- Custom autoscale

- select Custom autoscale

Choose how to scale your resource


**Manual scale**  
Maintain a fixed instance count


☐

**Custom autoscale**  
Scale on any schedule, based on any metrics or predictively


☒

Custom autoscale

Autoscale setting name	abcd-troweld-dev-vmss-Autoscale-297		
Resource group	abcd-troweld-dev-vmss_group		
Instance count ⓘ	3		
Predictive autoscale	Mode <div>Disabled ▾</div>	Pre-launch setup of instances (minutes) ⓘ <div></div>	
 Enable Forecast only or Predictive autoscale. <a href="#">Learn more about Predictive autoscale.</a>			

**Default\*** Auto created default scale condition 

Delete warning

 The very last or default recurrence rule cannot be deleted. Instead, you can disable autoscale to turn off autoscale.

Scale mode

☐ Scale based on a metric

☒ Scale to a specific instance count

Instance count\* ⓘ

2

Schedule

**This scale condition is executed when none of the other scale condition(s) match**

+ Add a scale condition

- select Scale based on a metric option

Save

Discard

Refresh

Logs

Feedback

Resource group

abcd-troweld-dev-vmss\_group

Instance count ⓘ

3

Predictive autoscale

Mode

Disabled

Pre-launch setup of instances (minutes) ⓘ

Enable Forecast only or Predictive autoscale.

[Learn more about Predictive autoscale.](#)

Default\*

Auto created default scale condition

Delete warning

The very last or default recurrence rule cannot be deleted. Instead, you can disable autoscale to turn off autoscale.

Scale mode

Scale based on a metric

Scale to a specific instance count

Rules

Scale out

When

abcd-troweld-dev-vmss

(Average) Percentag...

Incre:

Scale in

When

abcd-troweld-dev-vmss

(Average) Percentag...

Decre:

Instance limits

+ Add a rule

Minimum \* ⓘ

2

Maximum \* ⓘ

3

Default \* ⓘ

2

Schedule

This scale condition is executed when none of the other scale condition(s) match

+ Add a scale condition

- select Add Rule

Scale rule

Metric source

Current resource (abcd-troweld-dev-vmss)

Resource type

Virtual machine scale sets

Resource

abcd-troweld-dev-vmss

Criteria

Metric namespace \*

Virtual Machine Host

Metric name

Percentage CPU

1 minute time grain

Dimension Name	Operator	Dimension Values	Add
VMName	=	All values	+

If you select multiple values for a dimension, autoscale will aggregate the metric across the selected values, not evaluate the metric for each values individually.

80%

60%

40%

20%

0%

17:30

17:35

UTC+05:30

Percentage CPU (Average)

0.57 %

☐ Enable metric divide by instance count

Operator \*

Greater than

Metric threshold to trigger scale action \*

70

%

Duration (minutes) \*

10

Add

- here the default metric name is cpu percentage, let it be. based on the requirement you can choose from the list

20 / 36

- below it shows the Avg CPU usage

## Scale rule ✕

If you select multiple values for a dimension, autoscale will aggregate the metric across the selected values, not evaluate the metric for each values individually.

Percentage CPU (Average)

5.51 %

☐ Enable metric divide by instance count ⓘ

Operator \*

Greater than ▼

Metric threshold to trigger scale action \* ⓘ

70 %

Duration (minutes) \* ⓘ

5

Time grain (minutes) ⓘ

1

Time grain statistic \* ⓘ

Average ▼

Time aggregation \* ⓘ

Average ▼

Action

Operation \*

Increase count by ▼

Cool down (minutes) \* ⓘ

5

instance count \*

1 ✓

Update Delete

- Operator :
  - choose greater than
- Metric threshold to trigger scale action

- choose 70
- duration put 5 min [checks cpu for every 5 min]
- Operation :
  - select increase count by
- instance count = keep 1 [it means when cpu threshold increses by 70% then 1 vm will be created]

- select Add rule to add one more rule to decrease a vm if the CPU below 40%

5

g

card

①

cale

to cre

dition

## Scale rule

If you select multiple values for a dimension, autoscale will aggregate the metric across the selected values, not evaluate the metric for each values individually.

Percentage CPU (Average)

5.51 %

☐ Enable metric divide by instance count ⓘ

Operator \*  
Less than ▼

Metric threshold to trigger scale action \* ⓘ  
40 %

Duration (minutes) \* ⓘ  
5

Time grain (minutes) ⓘ  
1

Time grain statistic \* ⓘ  
Average ▼

Time aggregation \* ⓘ  
Average ▼

Action

Operation \*  
Decrease count by ▼

Cool down (minutes) \* ⓘ  
5

instance count \*  
1 ✓

Update

Delete

- save and come back

- below you can see instance limits

Save Discard Refresh Logs Feedback

Resource group abcd-troweld-dev-vmss\_group

Instance count ⓘ 3

Predictive autoscale Mode Disabled Pre-launch setup of instances (minutes) ⓘ

Enable Forecast only or Predictive autoscale. [Learn more about Predictive autoscale.](#)

**Default\*** Auto created default scale condition ⓘ

Delete warning ⓘ The very last or default recurrence rule cannot be deleted. Instead, you can disable autoscale to turn off autoscale.

Scale mode ☒ Scale based on a metric ☐ Scale to a specific instance count

Rules

Scale out

When	abcd-troweld-dev-vmss	(Average) Percentag...	Increase
[Progress bar]			

Scale in

When	abcd-troweld-dev-vmss	(Average) Percentag...	Decrease
[Progress bar]			

+ Add a rule

**Instance limits**

Minimum \* ⓘ 2 Maximum \* ⓘ 3

Default \* ⓘ 2

Schedule

**This scale condition is executed when none of the other scale condition(s) match**

+ Add a scale condition

- here you can see the instance count for min is 2 and max is 2 increase max by 1 that is 3, and default keep 2.
- Save the configuration and wait for few minutes, it will create one more vm as the stress increases to 100%.
- after the new vm is created and go to two machines and stop the execution.
- cmd to stop : ctrl+c
- wait for few minutes again it will delete one vm.



# Configure LoadBalancer for VMSS

A **load balancer** is a tool that distributes incoming traffic evenly across multiple servers to:

- Prevent overloading any single server.
- Ensure high availability and better performance.
- Automatically redirect traffic if a server goes down.

Think of it like a traffic cop directing cars (users) to different open lanes (servers) so everyone moves smoothly without jams.

## Layer 4 Load Balancer:

- Operates at the **transport layer** (Layer 4 of the OSI model).
  - Balances traffic based on **IP addresses** and **ports** (e.g., TCP/UDP traffic).
  - Does not inspect the actual content of the traffic (e.g., HTTP requests).
  - Faster and more efficient but less intelligent.
  - Example Use: Routing packets for a database or application server.
- 

## Layer 7 Load Balancer:

- Operates at the **application layer** (Layer 7 of the OSI model).
- Balances traffic based on **content of the request** (e.g., URL, headers, cookies).
- Can perform smart routing, such as sending requests for **/images** to one server and **/api** to another.
- More flexible but slightly slower due to deeper packet inspection.
- Example Use: Routing web application traffic (e.g., HTTP/HTTPS).
- Create Application gateway (layer 7 load balancer)/ load balancer
- Follow below images to create Application Gateway.

Home > Load balancing | Application Gateway >

# Create application gateway

- 1 Basics
- 2 Frontends
- 3 Backends
- 4 Configuration
- 5 Tags
- 6 Review + create

An application gateway is a web traffic load balancer that enables you to manage traffic to your web application. [Learn about creating application gateway](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* Azure subscription 1

Resource group \* vmssdemo

Create new

Instance details

Application gateway name \* appgateway

Region \* East US

Tier Standard V2

Enable autoscaling ☒ Yes ☐ No

Minimum instance count \* 1

Maximum instance count 5

Availability zone \* Zones 1, 2

IP address type ☒ IPv4 only ☐ Dual stack (IPv4 & IPv6)

HTTP2 ☐ Disabled ☒ Enabled

Configure virtual network

Virtual network \* Vnet1

Create new

Subnet \* appgate (10.0.1.0/24)

Manage subnet configuration

Home > Load balancing | Application Gateway >

# Create application gateway

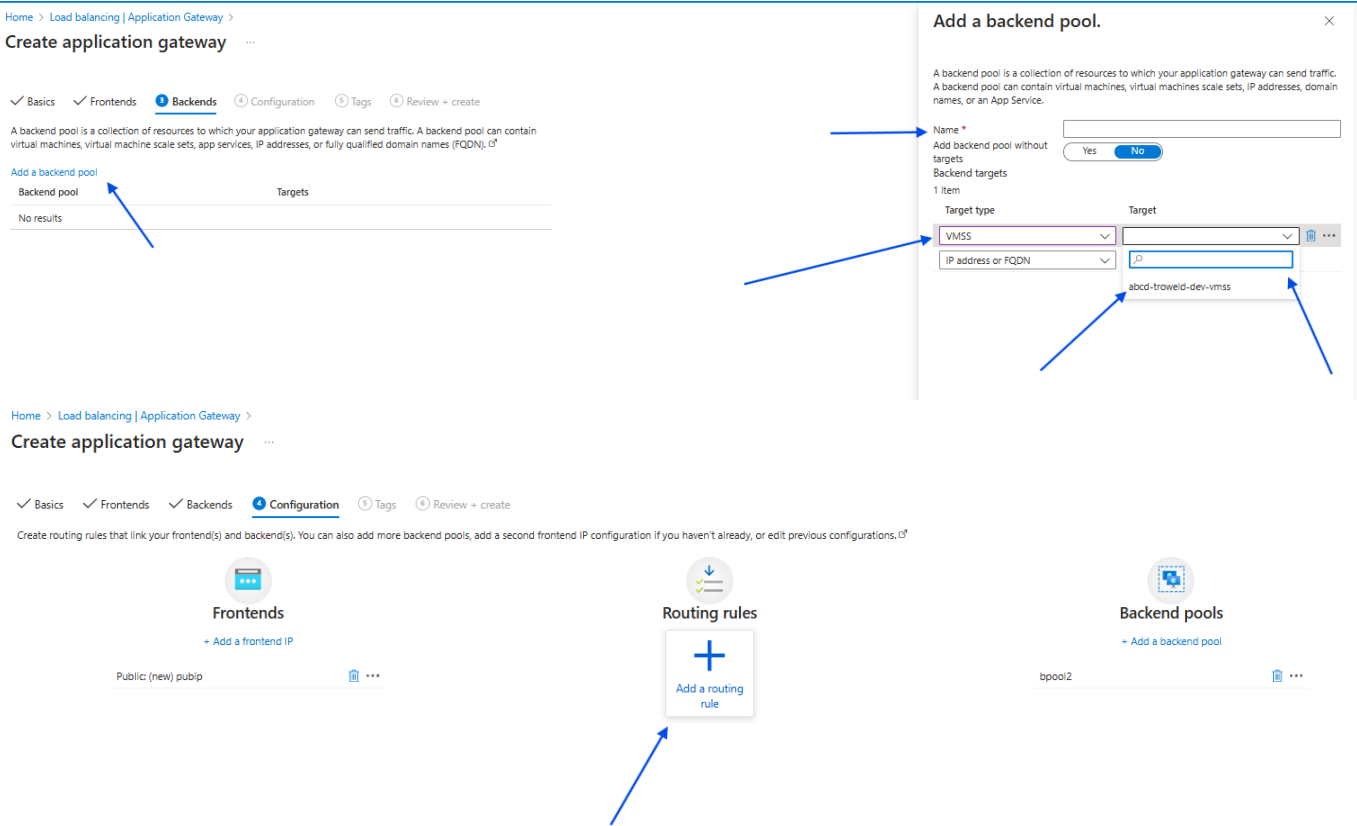
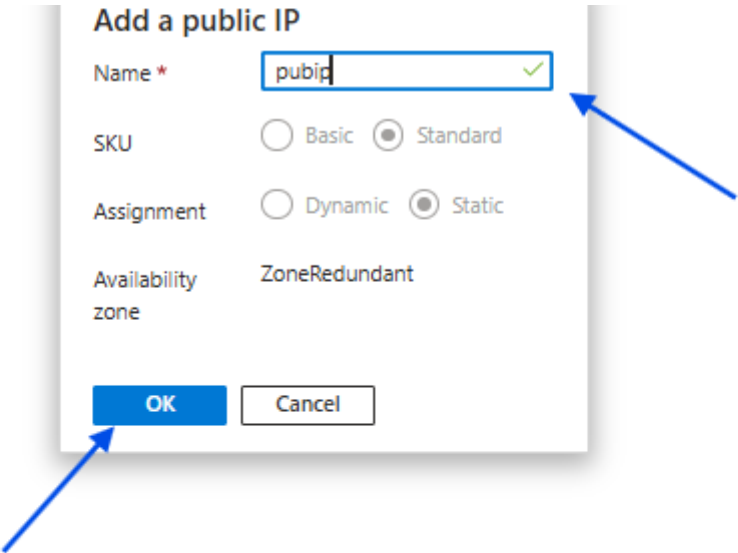
- ✓ Basics
- 2 Frontends
- 3 Backends
- 4 Configuration
- 5 Tags
- 6 Review + create

Traffic enters the application gateway via its frontend IP address(es). An application gateway can use a public IP address, private IP address, or one of each type.

Frontend IP address type ☒ Public ☐ Private ☐ Both

Public IPv4 address \* Choose public IP address

Add new



Routing rules

+

Add a routing rule

u haven't already, c

↓

✓

✓

×

Add a routing rule

Configure a routing rule to send traffic from a given frontend IP address to one or more backend targets. A routing rule must contain a listener and at least one backend target.

Rule name \*

rule1

✓

Priority \*

100

✓

\*Listener

\*Backend targets

A listener "listens" on a specified port and IP address for traffic that uses a specified protocol. If the listener criteria are met, the application gateway will apply this routing rule. ⓘ

Listener name \*

l1

✓

Frontend IP \*

Public IPv4

✓

Protocol ⓘ

☒ HTTP

☐ HTTPS

Port \*

80

✓

Listener type ⓘ

☒ Basic

☐ Multi site

Custom error pages

Show customized error pages for different response codes generated by Application Gateway. This section lets you configure Listener-specific error pages. [Learn more](#) ⓘ

Please verify that the url(s) being added here is reachable from your application gateway using the [connection troubleshoot](#) tool to prevent any deployment error.

Bad Gateway - 502

Enter Html file URL

Forbidden - 403

Enter Html file URL

[Show more status codes](#)

28 / 36

X



▼

▼

[Add new](#)

No additional targets to display

29 / 36

Home > Load balancing | Application Gateway >

## Create application gateway ...

✓ Validation passed

✓ Basics ✓ Frontends ✓ Backends ✓ Configuration ✓ Tags **6 Review + create**

### Basics

Subscription	Azure subscription 1
Resource group	vmssdemo
Name	appgateway
Region	East US
Tier	Standard_v2
Enable autoscaling	Enabled
Minimum instance count	1
Maximum instance count	5
Availability zone	Zones 1, 2
HTTP2	Enabled
Virtual network	Vnet1
Subnet	appgate (10.0.1.0/24)
Subnet address space	10.0.1.0/24

### Frontends

Public IPv4 address name	pubip
SKU	Standard
Assignment	Static
Availability zone	ZoneRedundant

### Tags

None

Create

Previous

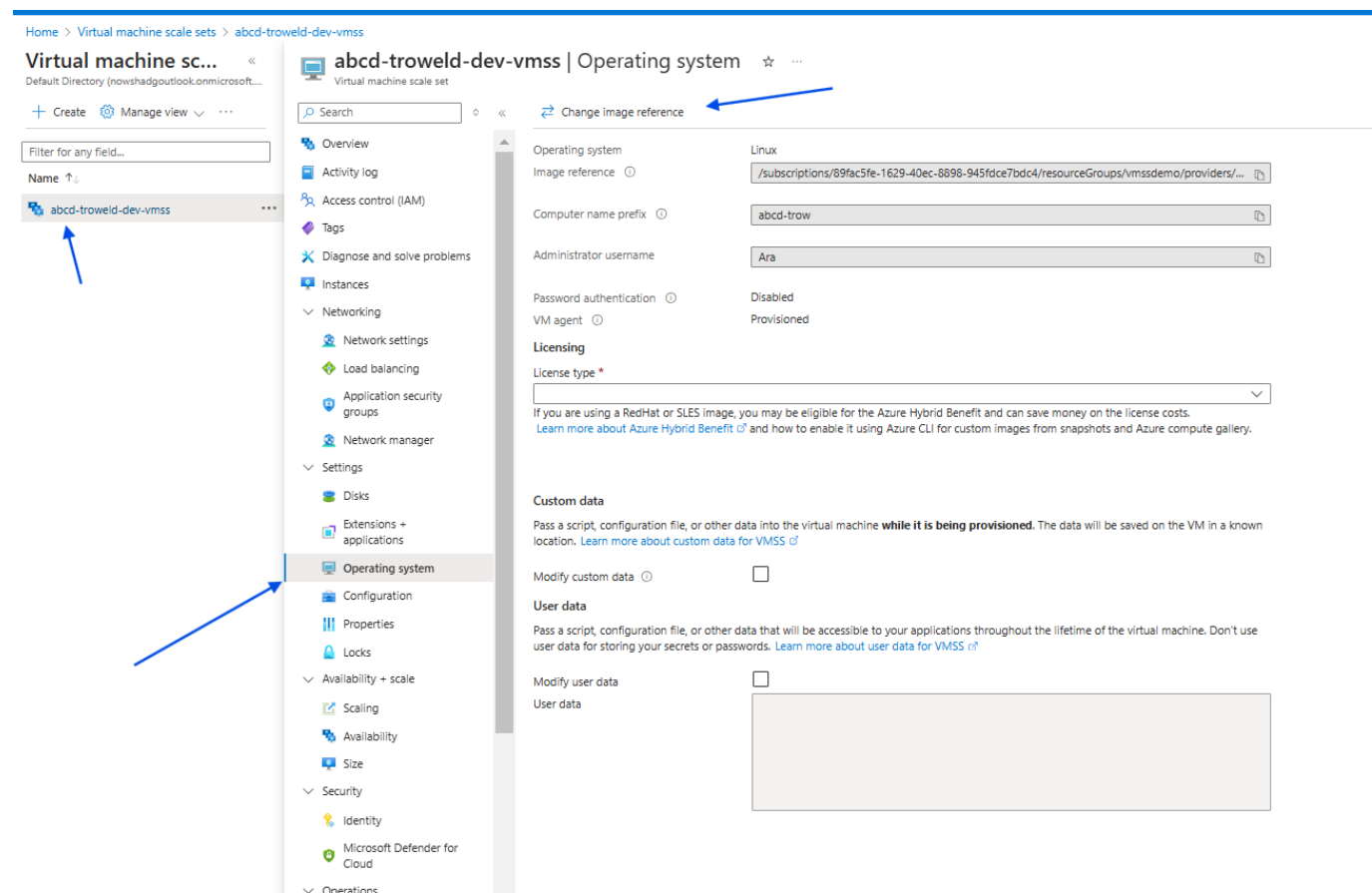
Next

[Download a template for automation](#)

- Take the Loadbalenser ip and check whether the site is working or not
- if it is working then you have successfully created the load balancer for vmss.

## Change image for VMSS [upgrading the image]

- Create a new VM and deploy different website and take image from it. and give version 1.1.0.



- Open the VMSS under, see above image select **Operating System** and top left side you can see **Change image reference**

Home > Virtual machine scale sets > abcd-troweld-dev-vmss | Operating system >

Change image reference

You can update to a newer VM image version or to a new image altogether. Migrating to Azure compute gallery provides benefits like easy updates to the latest version when available and global replication so that the same image can be used across regions.

Select image source

☒ Azure compute gallery

☐ Platform image

Azure compute gallery

Webserver

VM image definition

Troweld\_Site

VM image version

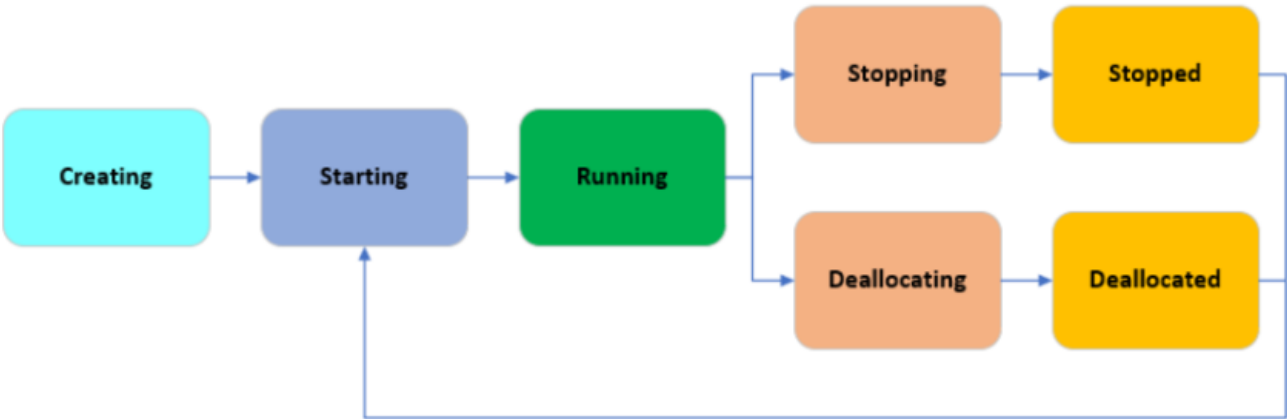
Select a VM image version

1.0.0

1.1.0

- change the image by seeing above image
- Apply
- The upgrade will be done automatically after few minutes
- check manually and through load balancer if it comes with new site the deployment is succesful.

Azure VM States





- Above images shows the multiple steps a vm undergoes during creation. [Azure official docs](#)

 Expand table

Power state	Description	Billing
Creating	Virtual machine is allocating resources.	Not Billed*
Starting	Virtual machine is powering up.	Billed
Running	Virtual machine is fully up. This state is the standard working state.	Billed
Stopping	This state is transitional between running and stopped.	Billed
Stopped	The virtual machine is allocated on a host but not running. Also called <i>PoweredOff</i> state or <i>Stopped (Allocated)</i> . This state can be result of invoking the <code>PowerOff</code> API operation or invoking shutdown from within the guest OS. The <i>Stopped</i> state might also be observed briefly during VM creation or while starting a VM from <i>Stopped (Deallocated)</i> state.	Billed
Deallocating	This state is transitional between <i>Running</i> and <i>Deallocated</i> .	Not billed*
Deallocated	The virtual machine has released the lease on the underlying hardware. If the machine is powered off it is shown as <i>Stopped (Deallocated)</i> . If it has entered <a href="#">hibernation</a> it is shown as <i>Hibernated (Deallocated)</i>	Not billed*

- the stop which we do from portal leads to deallocation, so no bill. but there will disk charges

## Disk Related Options in Azure VM

- In Azure we have virtual disks managed by Managed Disk and Temp Disk (Local storage)
- Data in local storage will be erased when you stop instance
- Azure uses Data Disks (Managed disks) for persistence even after vm is stopped or deleted
- In Azure many instance sizes come with local storage and information is present in vm size detailing
- Azure has a restriction on number of disks according to vm size

vm1 | Size

Search

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Networking

Connect

Disks

Size

Microsoft Defender for Cloud

Advisor recommendations

Extensions + applications

Availability + scaling

Operating system

Configuration

Identity

Properties

Locks

Operations

Monitoring

Automation

Help

If the virtual machine is currently running, changing its size will cause it to be restarted. Stopping the virtual machine may reveal additional sizes. →

Search by VM size...

VCPUs: All

RAM (GiB): All

Display cost: Hourly

Add filter

Showing 483 VM sizes. | Subscription: Azure subscription 1 | Region: East US | Current size: Standard\_B1s | [Learn more about VM sizes](#) | 

Group by series

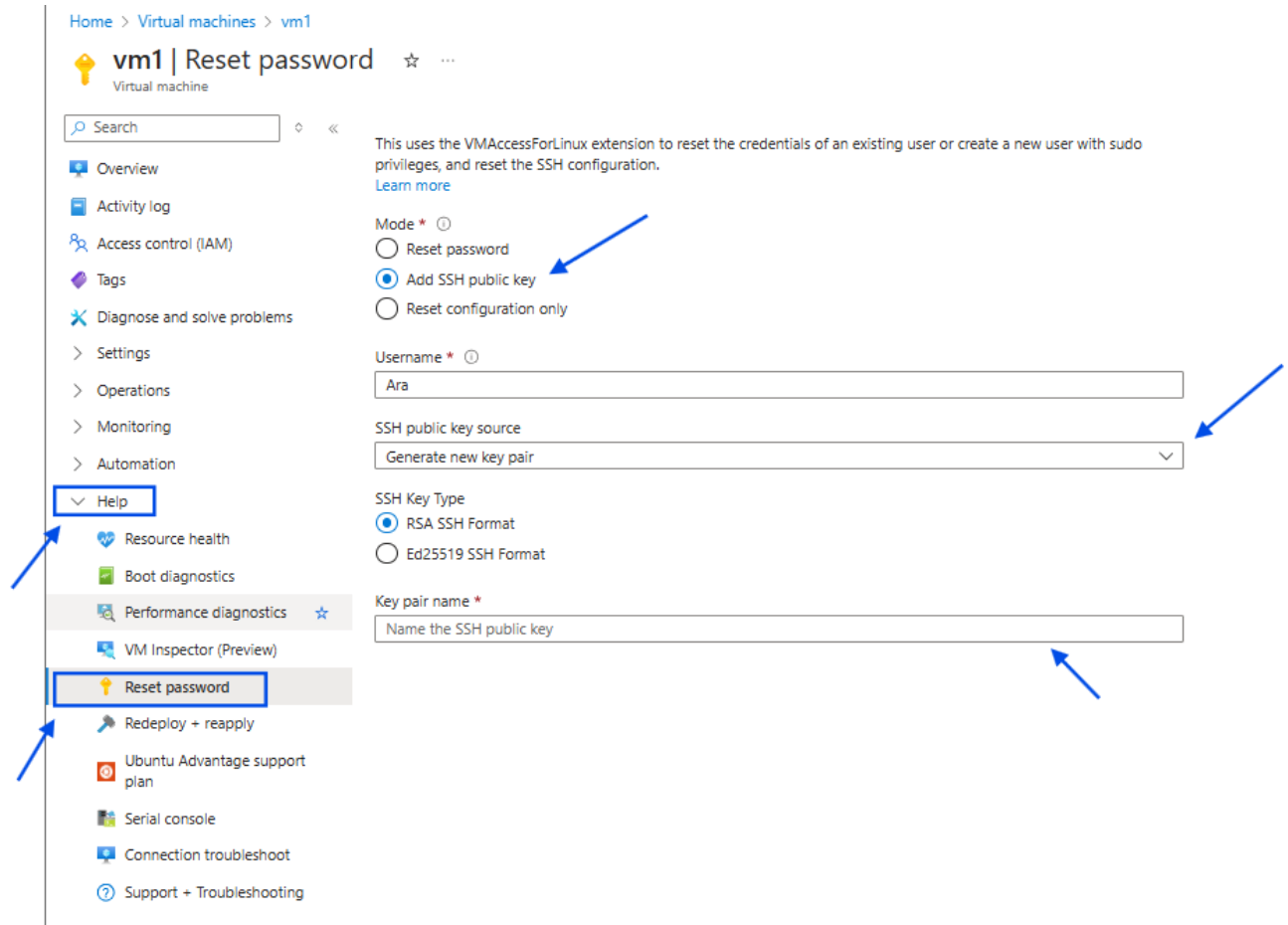
VM Size	Type	VCPUs	RAM (GiB)	Data disks	Max IOPS	Local storage (GiB)	Premium disk	Cost/hour
Most used by Azure users								
The most used sizes by users in Azure								
DS1_v2	General purpose	1	3.5	4	3200	7 (SCSI)	Supported	\$6.07
D2s_v3	General purpose	2	8	4	3200	16 (SCSI)	Supported	\$7.99
B2s	General purpose	2	4	4	1280	8 (SCSI)	Supported	\$3.46
B1s (free services eligible)	General purpose	1	1	2	320	4 (SCSI)	Supported	\$0.87
B2ms	General purpose	2	8	4	1920	16 (SCSI)	Supported	\$6.92
B1ms	General purpose	1	0.5	2	320	4 (SCSI)	Supported	\$0.43
DS2_v2	General purpose	2	7	8	6400	14 (SCSI)	Supported	\$12.15
B4ms	General purpose	4	16	8	2880	32 (SCSI)	Supported	\$13.81
D4s_v3	General purpose	4	16	8	6400	32 (SCSI)	Supported	\$15.97
DS3_v2	General purpose	4	14	16	12800	28 (SCSI)	Supported	\$24.38
DB3_v3	General purpose	8	32	16	12800	64 (SCSI)	Supported	\$31.95
D-Series v4								
The 4th generation D family sizes for your general purpose needs								
Ideal for workloads that do not need continuous full CPU performance								
B2s	General purpose	2	4	4	1280	8 (SCSI)	Supported	\$3.46
B1s (free services eligible)	General purpose	1	1	2	320	4 (SCSI)	Supported	\$0.87
B2ms	General purpose	2	8	4	1920	16 (SCSI)	Supported	\$6.92
B1ms	General purpose	1	0.5	2	320	4 (SCSI)	Supported	\$0.43
B4ms	General purpose	4	16	8	2880	32 (SCSI)	Supported	\$13.81
B1ms	General purpose	1	2	2	640	4 (SCSI)	Supported	\$1.72
B8ms	General purpose	8	32	16	4320	64 (SCSI)	Supported	\$27.70
E-Series v4								
The 4th generation E family sizes for your high memory needs								

## Agents in Azure VM

- Azure VM will have an inbuilt agent running inside VMs
  - Linux VM Agents [azure official docs](#)
  - Windows VM Agents [azure official docs](#)

### Activity : LOST PEM File in Azure

- Open the VM in azure.
- from the left side blade select click on **Help**, from the list select **Reset Password**

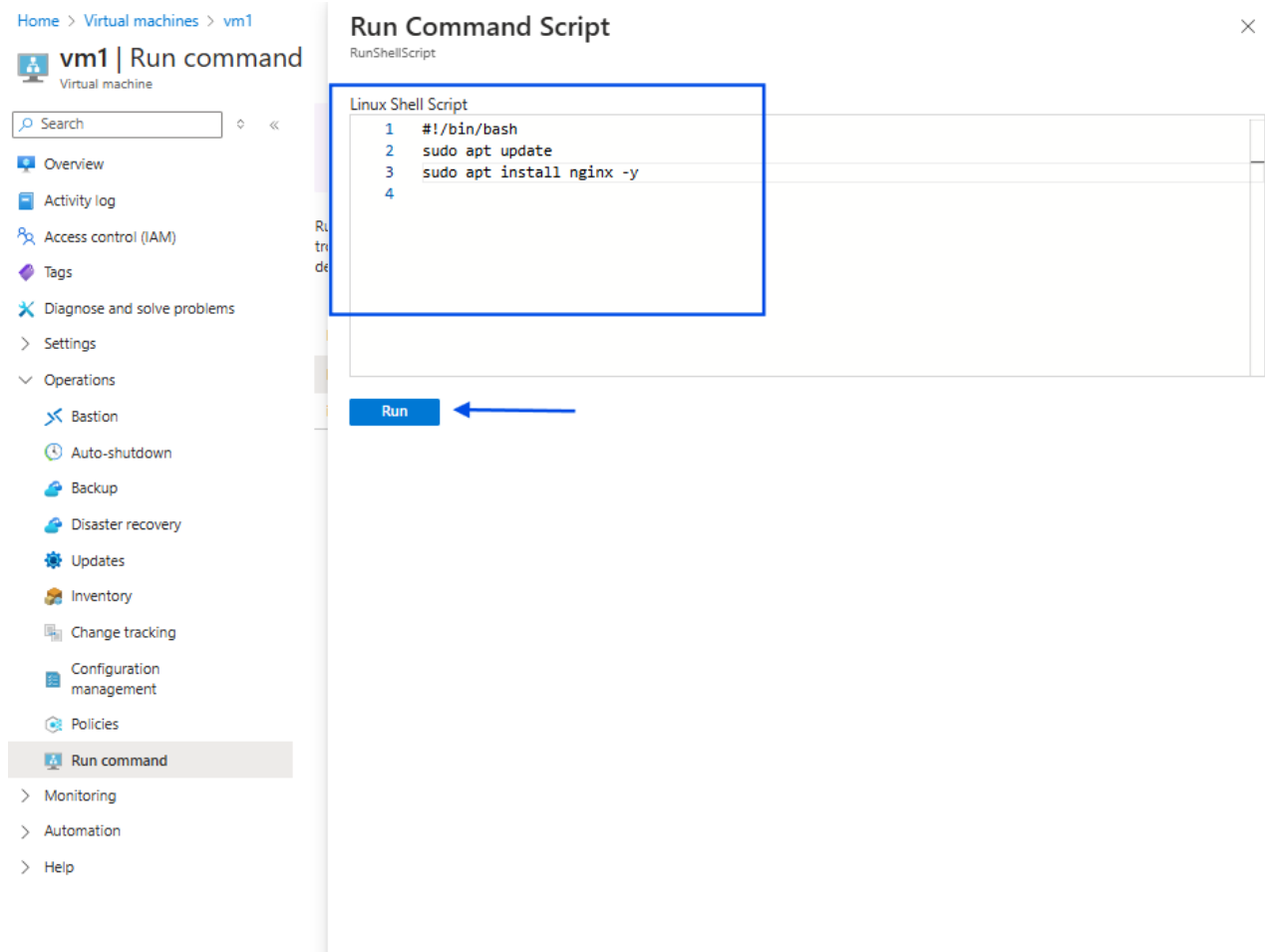
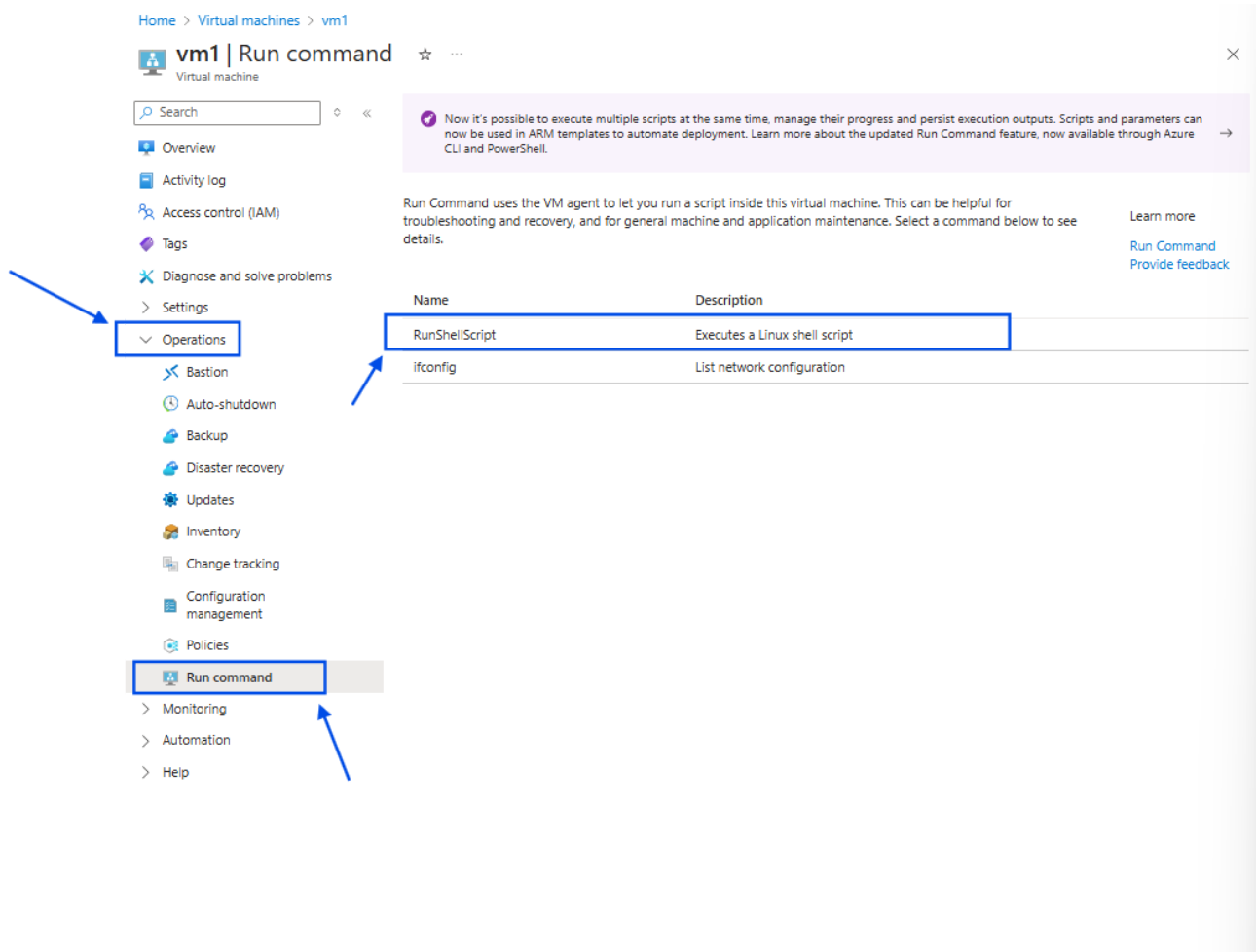


- fill the necessary details
  - create a new SSH key from **SSH key** services
  - select **Add SSH public Key**
  - let username be same
  - select **Use existing key stored in Azure**
  - now select the new key and **Update** the change.

if you want to Run any script without logging into Machine even after lost the pem file

- Open the VM ,find the **Operations** section expand it.
- find **Run Command** option and click on it.

- select **RunShellScript**



- enter above script and run, it will execute the script and shows output below

- to test, take public ip and browse it you should get nginx page where a sign of script executed correctly.