DEPLOY A VIRTUAL MACHINE SCALE SET BEHIND A LOAD BALANCER

Scale set

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.

We need to create load balancer before creating a scale set, make sure both the load balancer and scale set are in the same vnet

To create a load balancer,

- Portal > search load balancer > create > enter the following details
- In basics tab,

SETTING	VALUE
Subscription	Select your subscription.
Resource group	demo-RG
Name	demo-load-balancer
Region	Central India
SKU	Standard
Tier	Regional

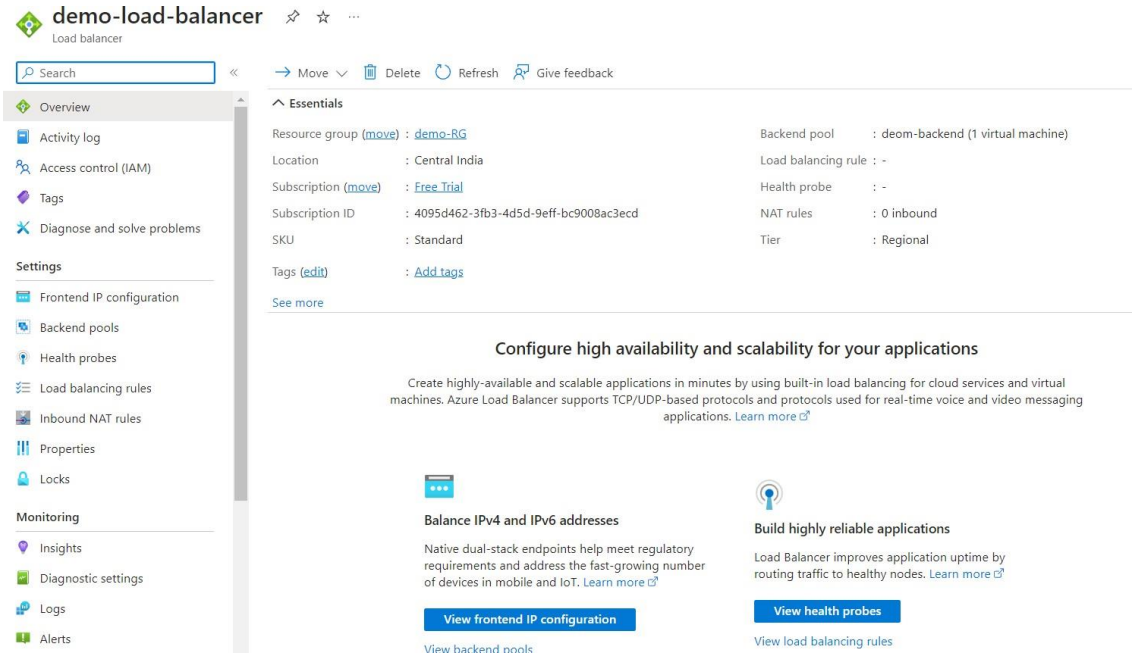
- In frontend IP configuration > add IP

SETTING	VALUE
Name	demo-fip
Vnet	demo-Vnet
subnet	dev4

- In backend pools,

SETTING	VALUE
Name	demo-backend
Vnet	demo-Vnet
add	vm-1

- Select review + create > create



To create a scale set,

- Portal > search scale set > create
- On the basics tab, enter the following:

Setting	Value
Subscription	Select your subscription.
Resource group	demo-RG
vm Scale set name	demp-scaleset
Region	Central india
Availability	none
orchestration mode	flexible
security type	standard
image	windows server 2019
size	standard_B1s

- After entering username and password, on the scaling tab enter the following

Initial instance count * 1

Scaling

Scaling policy
- Manual
- Custom

Minimum number of instances * 1

Maximum number of instances * 2

Scale out

CPU threshold (%) * 2

Duration in minutes * 5

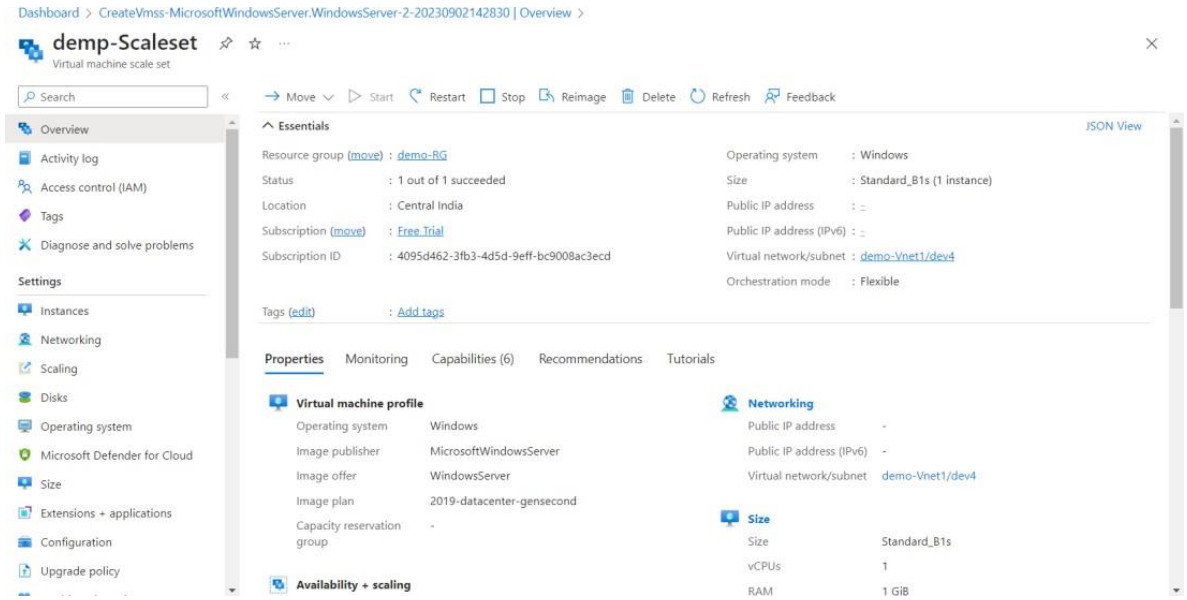
Number of instances to increase by * 1

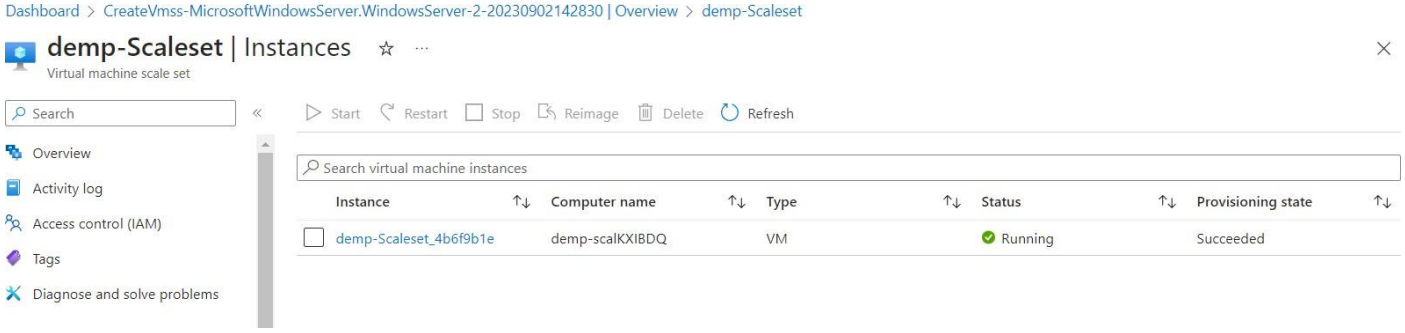
Scale in

CPU threshold (%) * 1

Number of instances to decrease by * 1

- In the networking tab, select the appropriate subnet where the previous load balancer is created
- Enter review + create > create





Working

We can observe that a virtual machine is being created automatically when the cpu threshold is reached