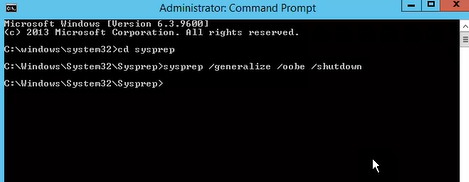
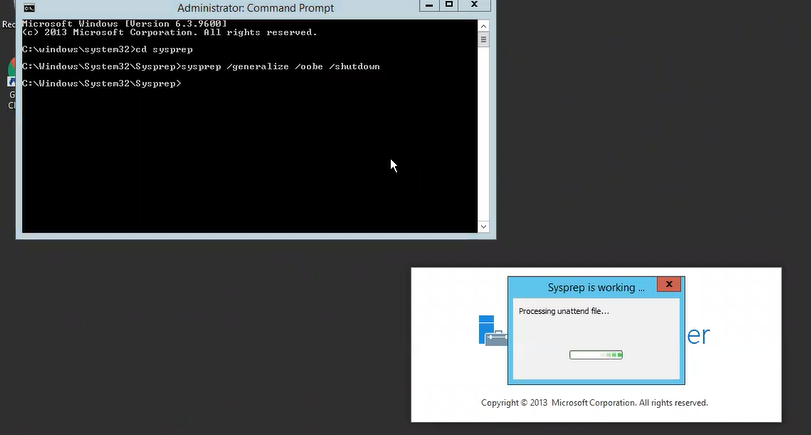
Capturing

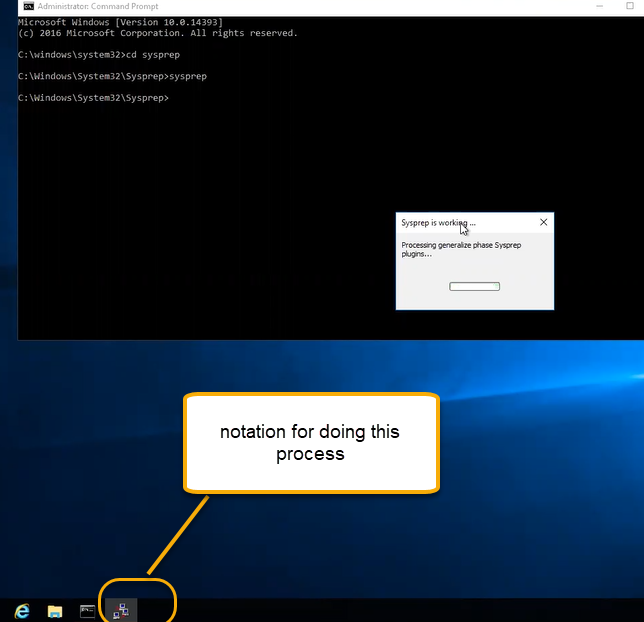




We can run the sysprep only once another time if you run it will not work

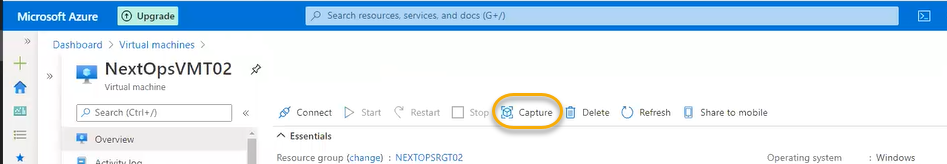
If the error will occurs then it might be the image (O.S) error or underline server issue

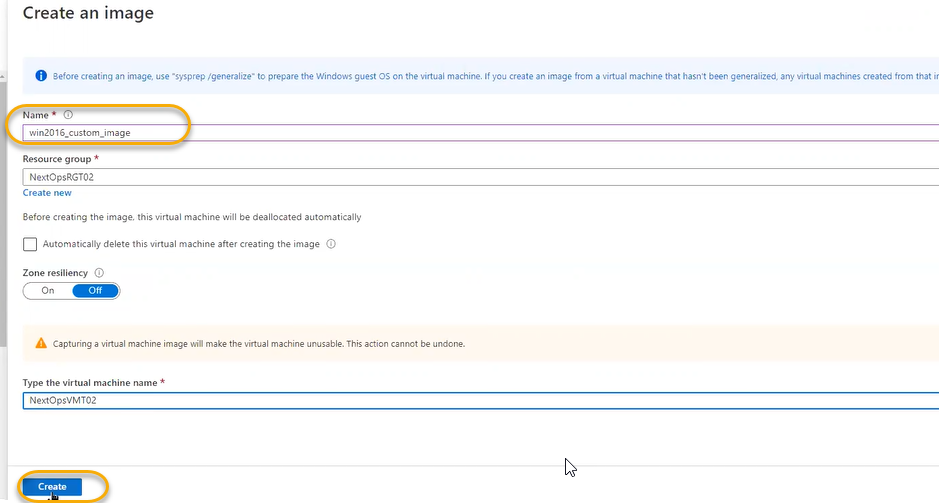
The sysprep option is only windows machine and the image created for the O.S will be sysprep by windows



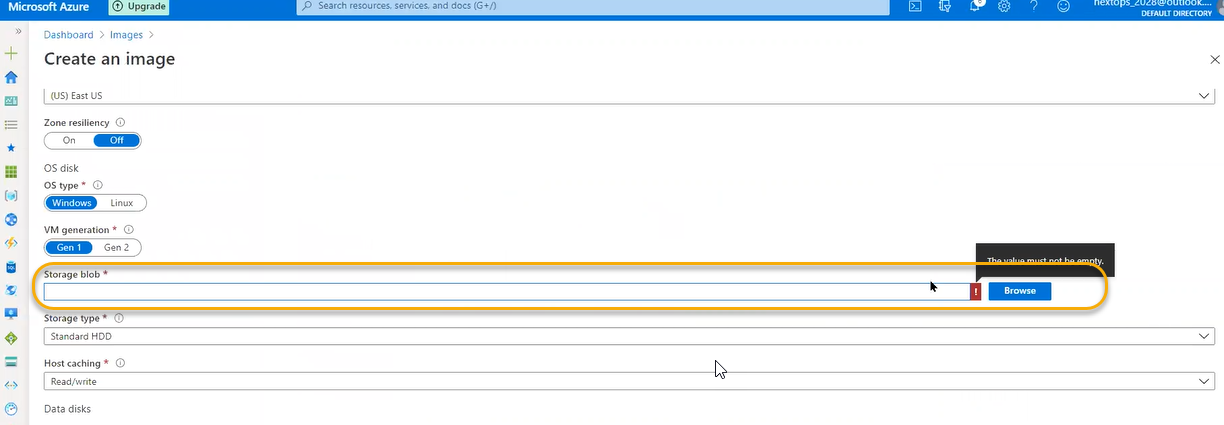
Once the system is down then stop the VM

Click in the capture



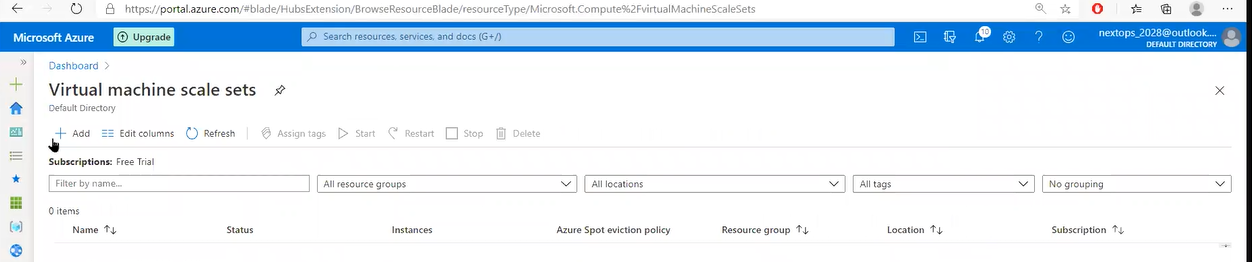


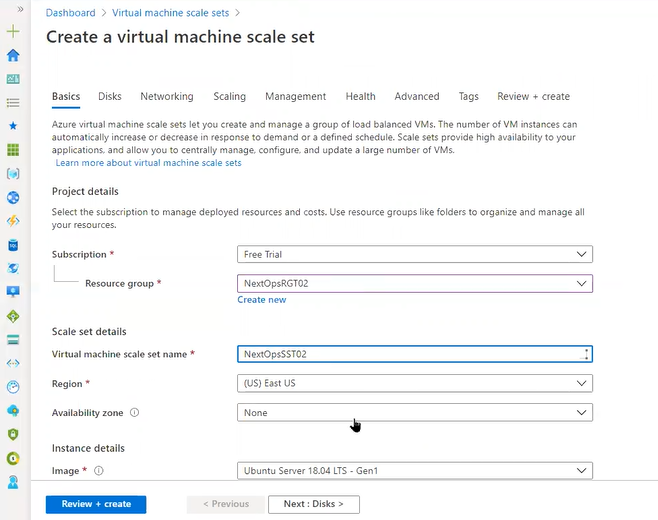
For importing the image from on-premises the file will be stored in storage accounts and while creating a image we can select the option as highlighted below and select the image from storage account

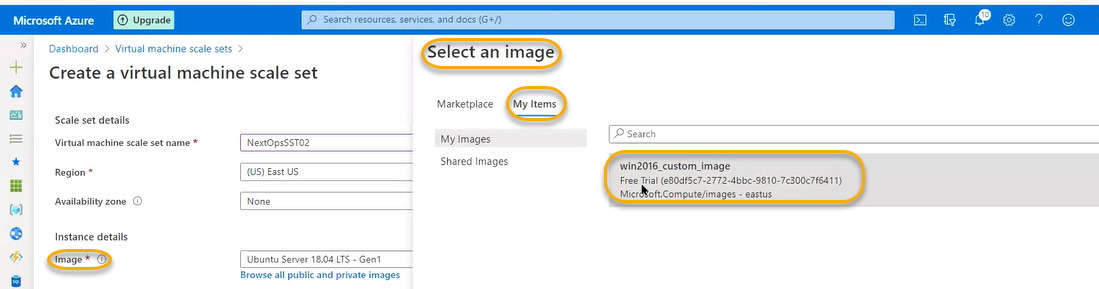


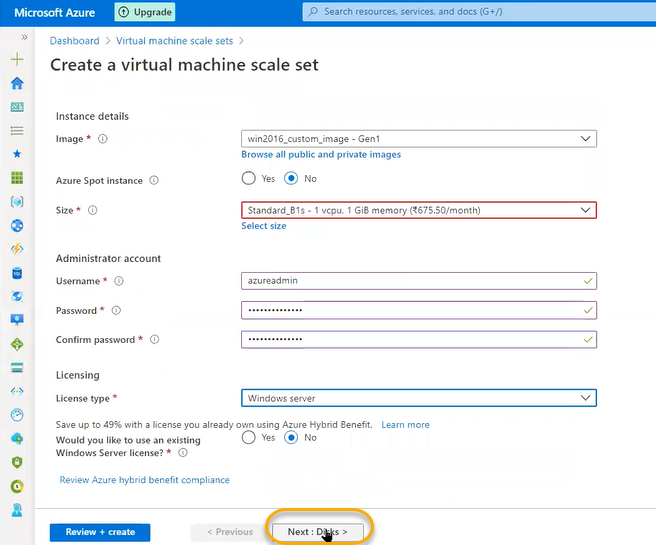
We don’t need to run the sysprep

Since we have created the image the next step is create the virtual machine scale sets



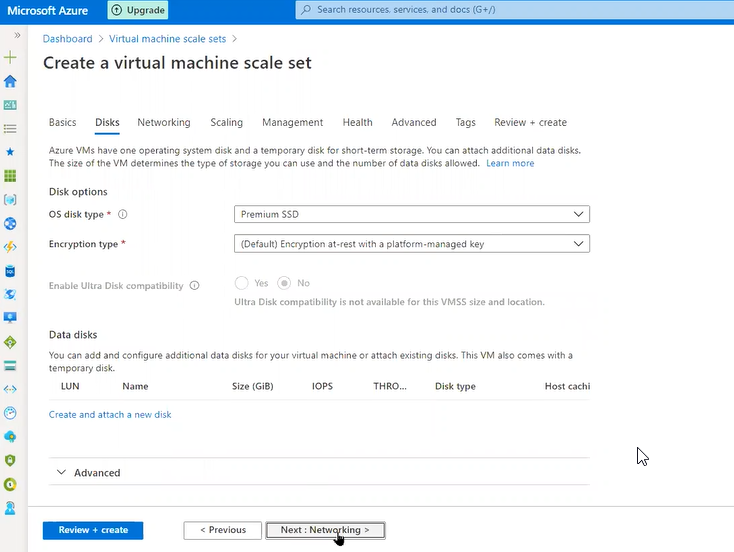


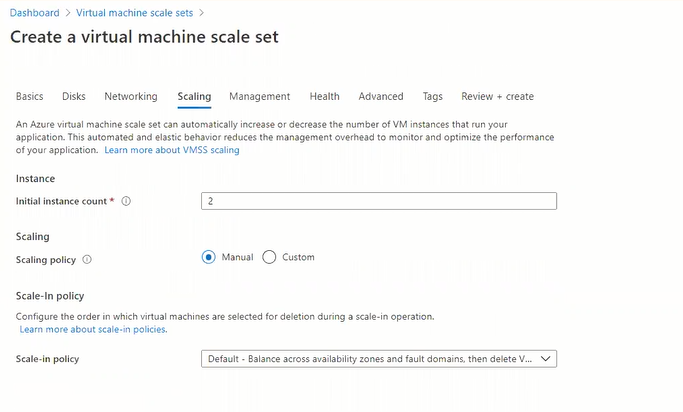




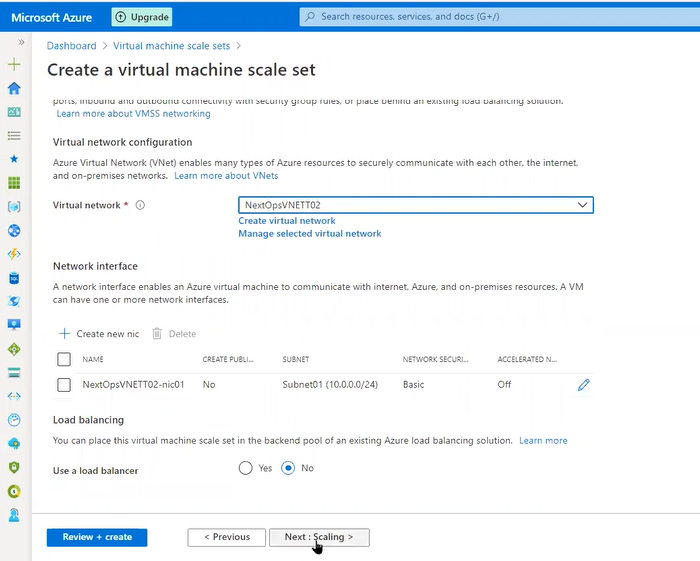
License type is windows server

There is option in azure called windows virtual desktop (VDI) in that only windows 10 operating systems will be there 2016 and 2012 O.S and kernel level is same whereas for the 2008 and 2006 have different O.S and kernel





We have given initial count as 2 so only 2 scale set are created



Select the default options and click and next and review and create

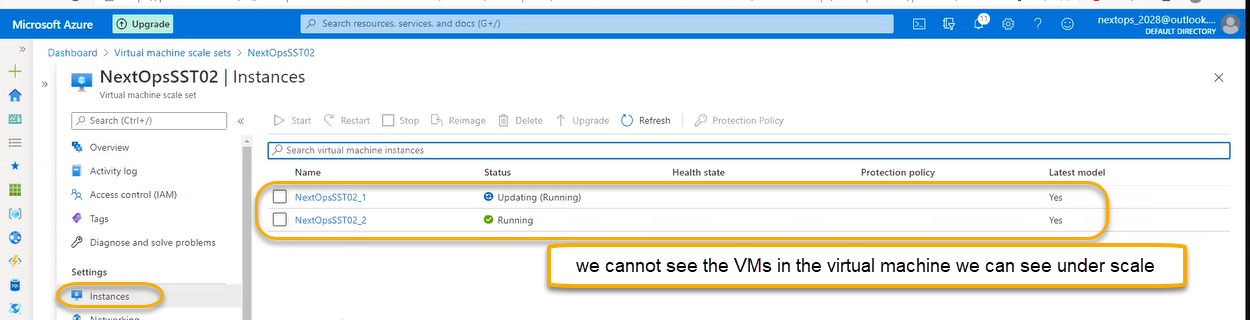
Scale set is automatic expansion and contraction of infrastructure based on traffic and it is always placed in load balancer so that based on load automatic expansion or contraction will appear

Machines are collectively working we need to bring the single identification and all of them are mirror images all the machines have installed web server in that scenario if end user send the request it will not go to specific to one instance

In the scale set there are 10 instances for 10 instances there are 10 names if the end user is not accessing the URL if they want to access the website and they want the same experience first we need mirroring for all the same instance we have same content

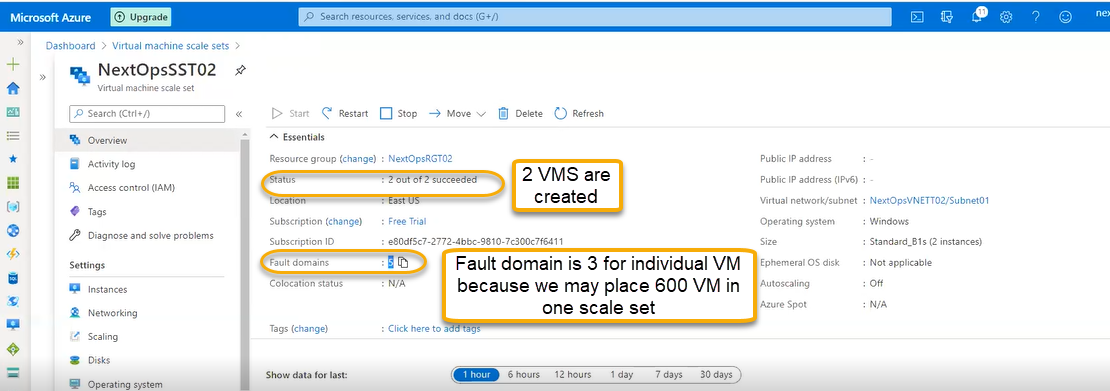
Instead of using load balancing if you want to use mirror for the 10 instances, we have 10 DNS and that 10 DNS we need to give to the users to access the website ( which is a complex process )

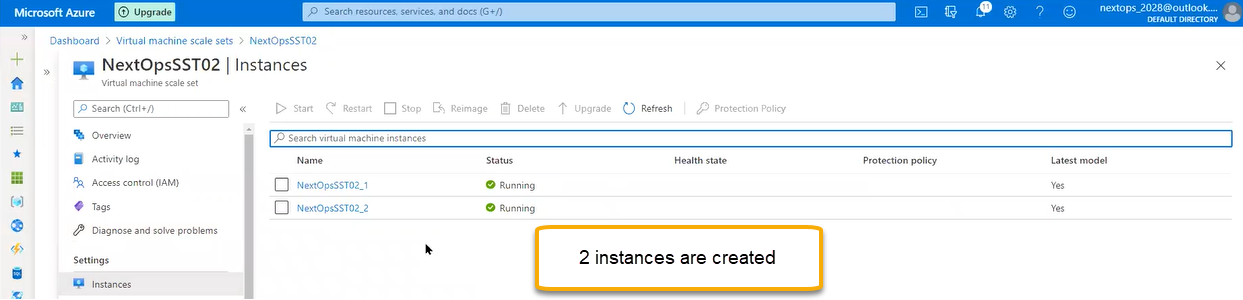
So for the better experience User will hit the load balancer and then request will go to the any of the user interference

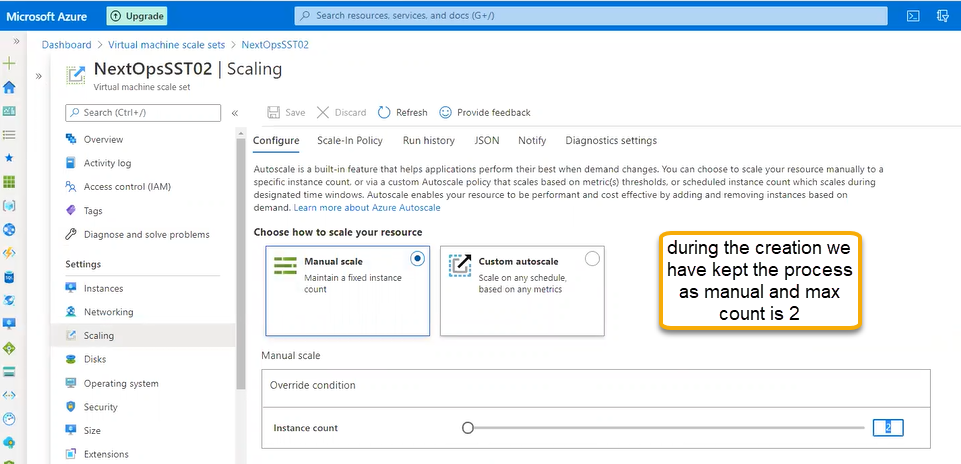


VM does not have much options instead we can see them in the scale sets

If you are doing for manual we need to hard coding or automatic means we can configure threshold





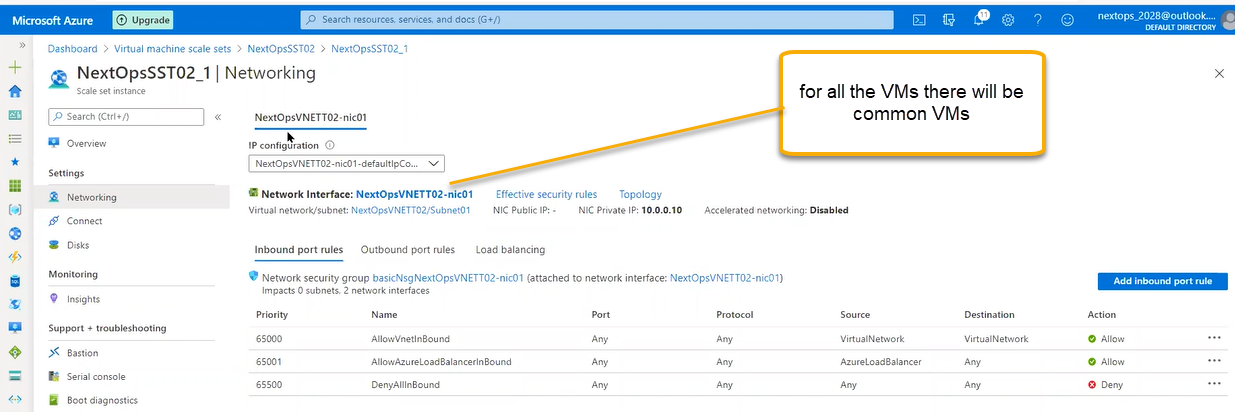


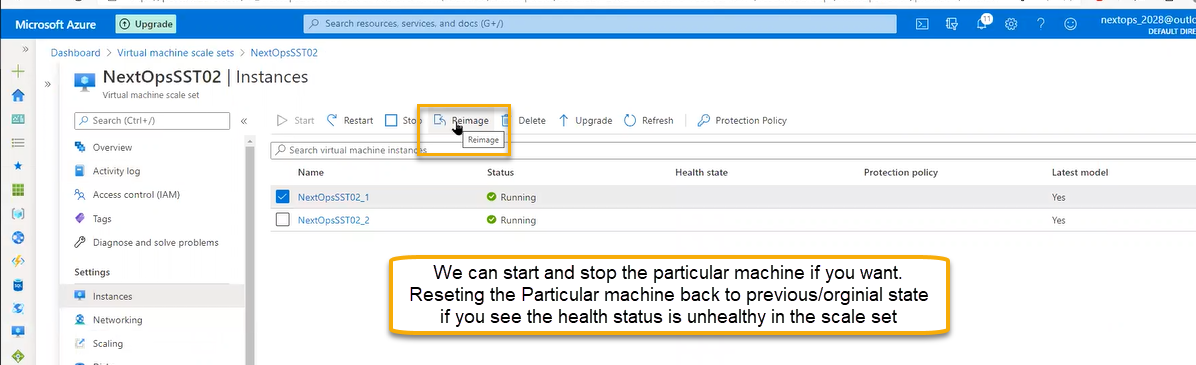
In the manual there is no algorithm or logic there is no expansion and contraction of instances in the manual scaling we can extend upto 600 VMS

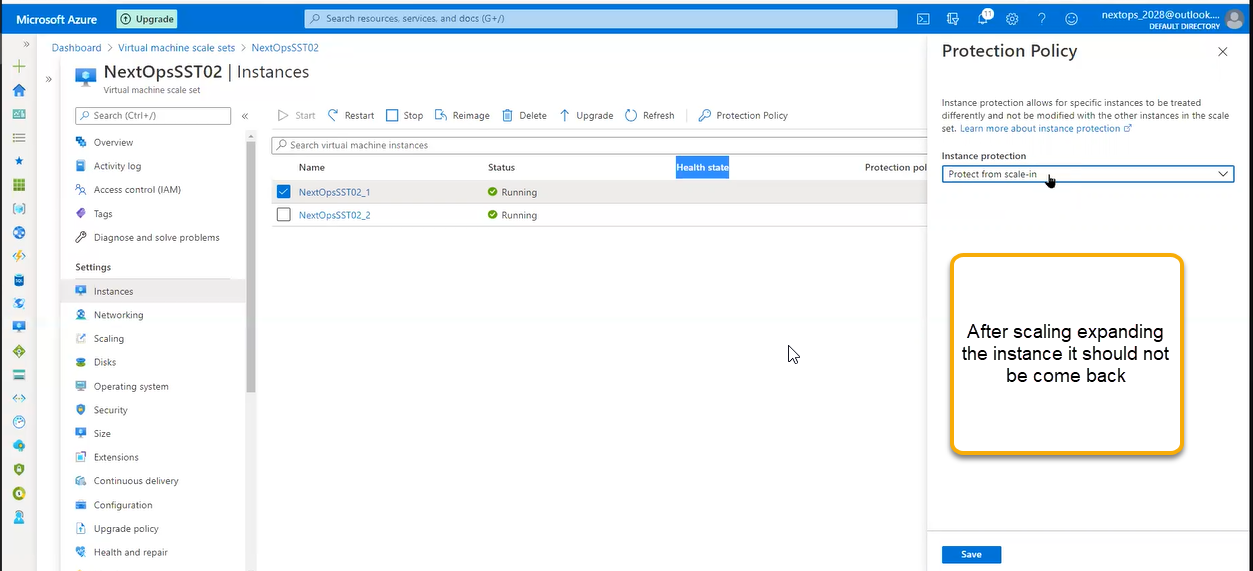
Cost will be increased in manual VMS



Based on instance ID we can scale back the VMS







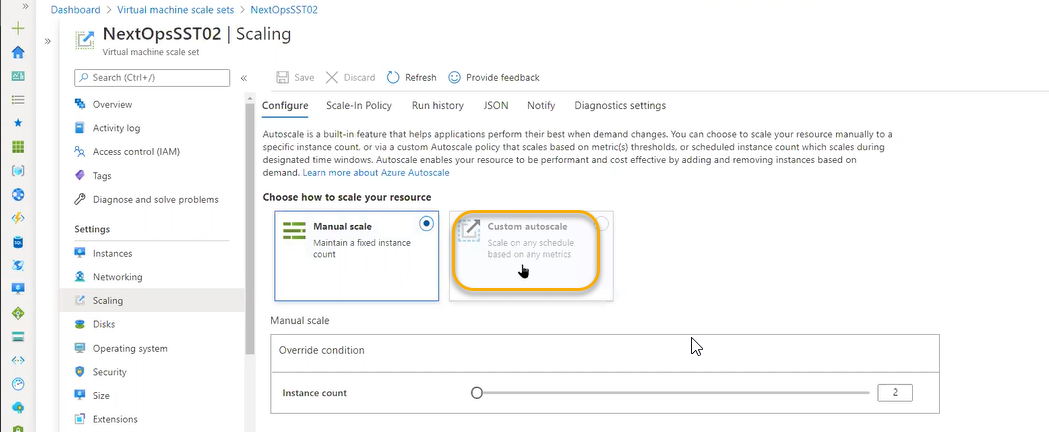
Scaling

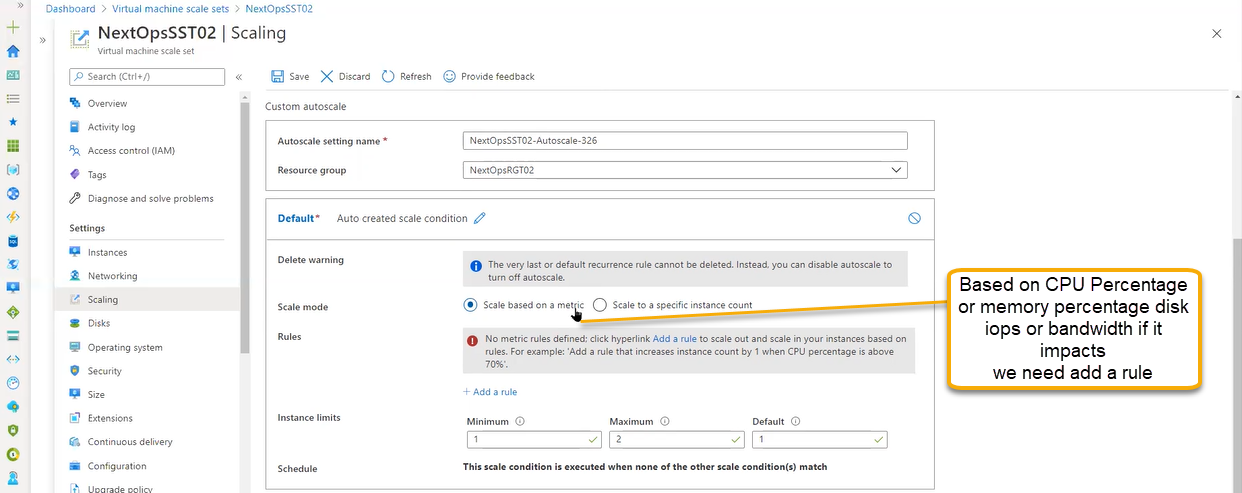
Manual scaling:it is hard coded value and the machines will be always running here there is NO Scale In and Out

If you reduce instant count as 1 and save it then once instance will be dead

Based on the load we have to do manual intervention

Auto scaling: Purpose of scaling is doing the Auto Scaling method





Click on Add a rule as shown in below screenshot the scale rule will pop up



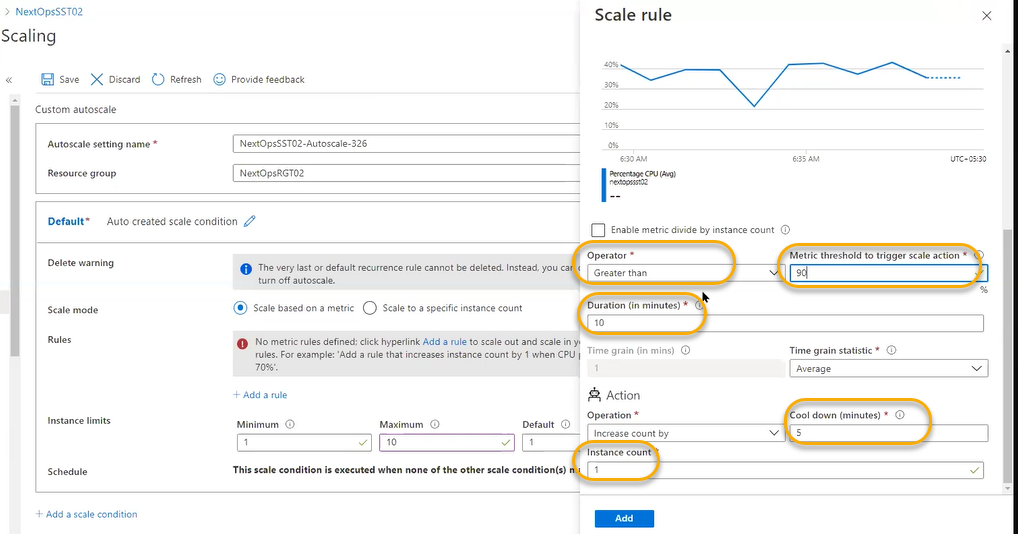
If the 1st instance CPU is more than 90 percent continuously in 10 minutes ( initially one instance is created when ever demand will be increased ) increase count by 1 instance immediately scale set will deploy another instance

sometimes there will be false positives the load is NOT generated by end user may be antivirus program will be scanning or any back up will be running or Java Ex will be running in that cases there will be false positives or some operating process running in the O.S in that cases we need to make sure to take correct decisions like monitoring solutions, performance diagnostics

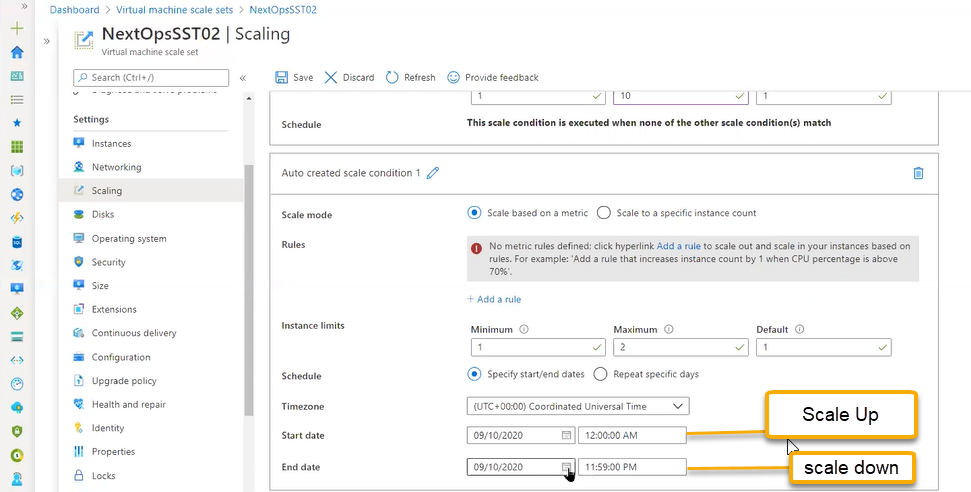
or if we map monitoring to scale set, we can take decision

in that cases we need to postpone our updates and the issue is causing due to Java Ex that is a bug we need to uninstall

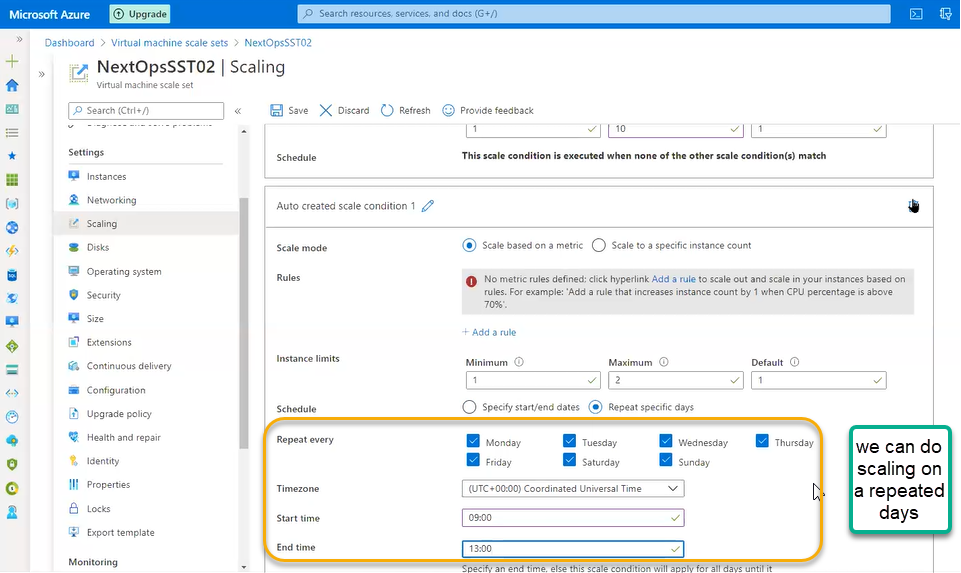
cooldown is nothing but once the instance is created it will take 5 minutes for UP

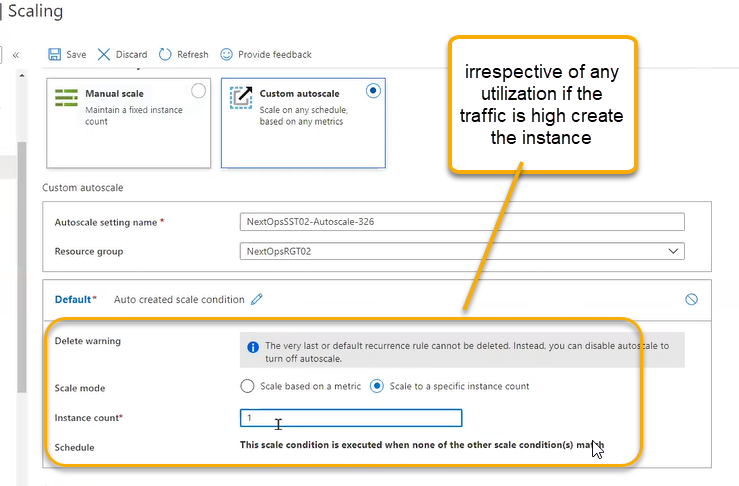


Another scale conditions will be scheduled based for specific time instances has to be created or decreased irrespective of CPU Utilization

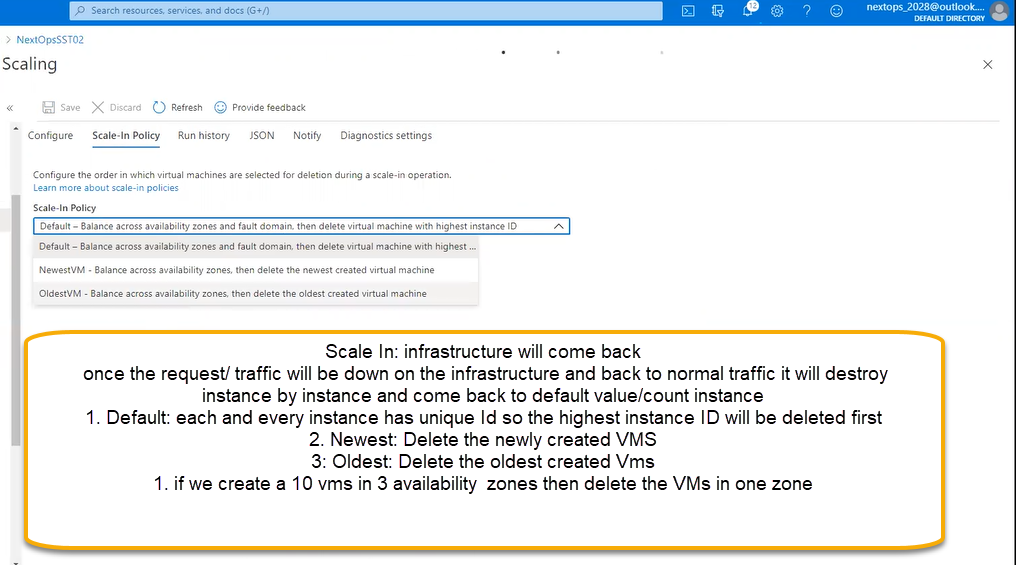


And we need to save it





Custom auto scale is the recommended option



1. Default:

Availibity Zone

1 2 3

VMS Placement

3 3 3

if traffic is low

then delete one VM in One Zone

Availibity Zone

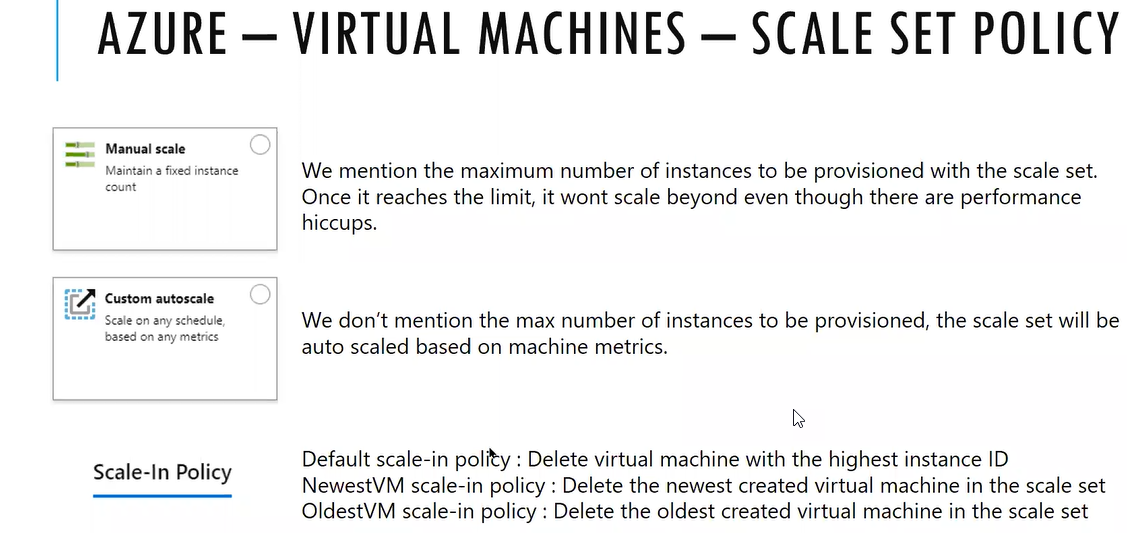
1 2 3

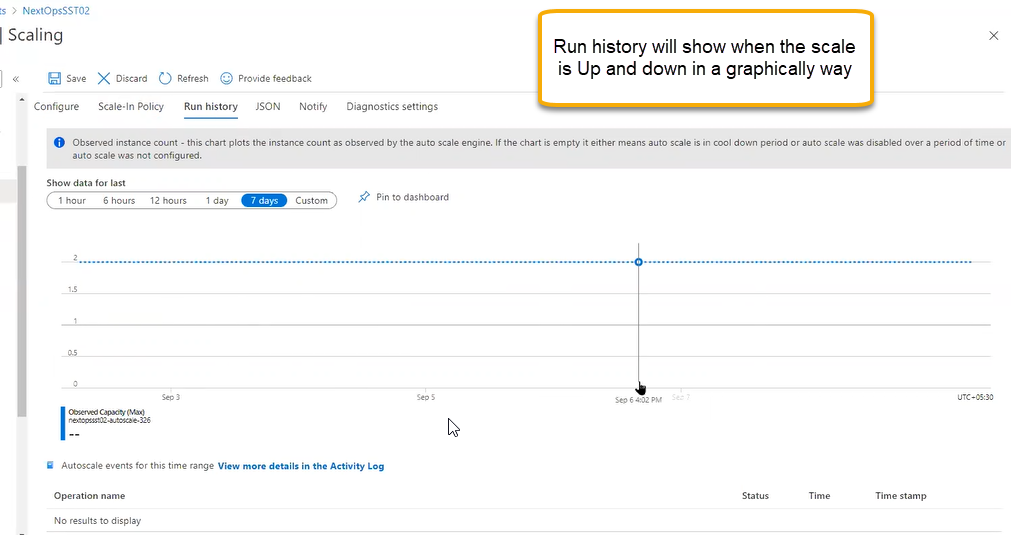
VMS Placement

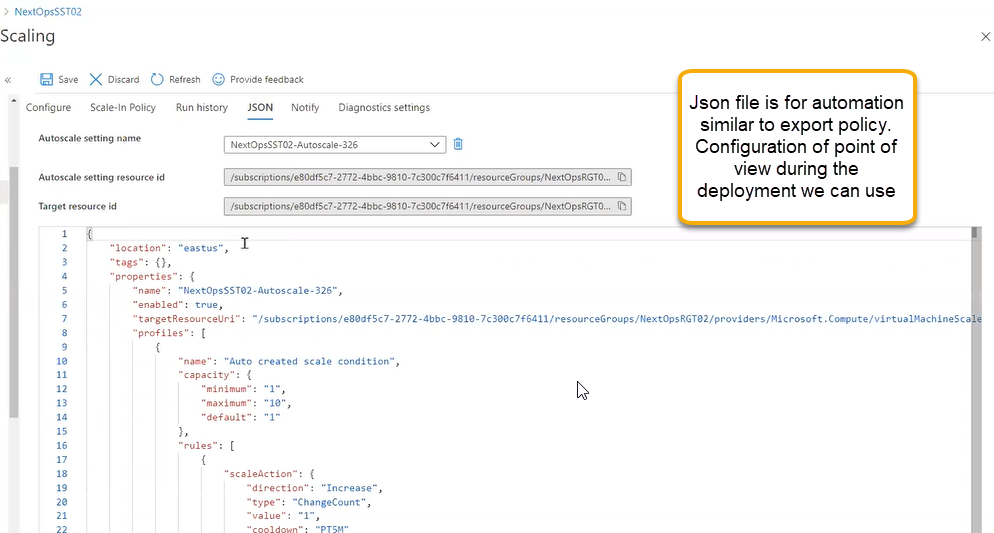
2 2 2

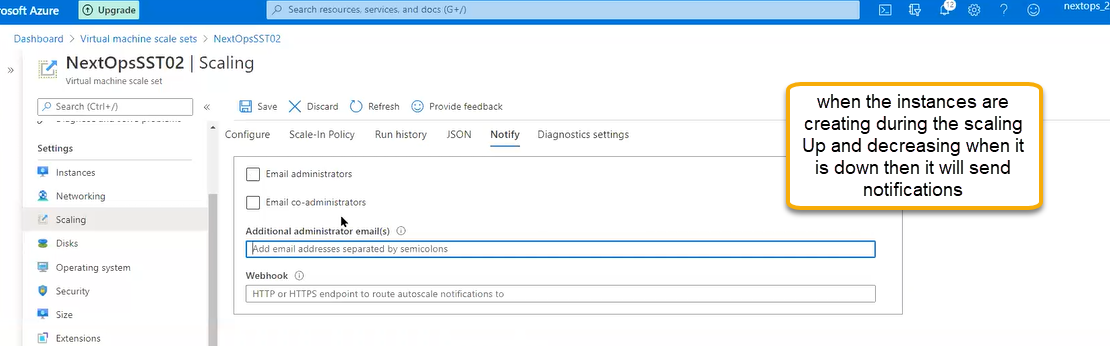
and so it will decrease VMs zone by zone

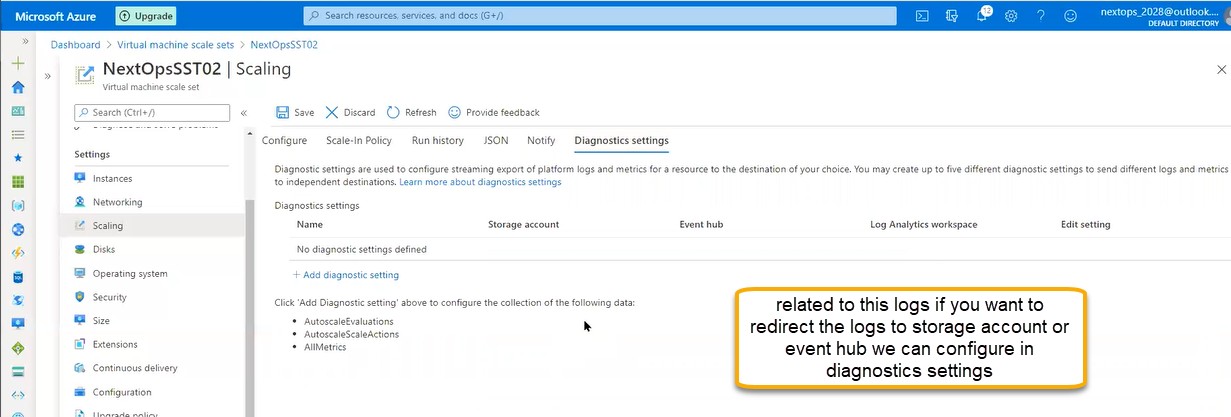
1. Scale in process is applicable to auto scaling

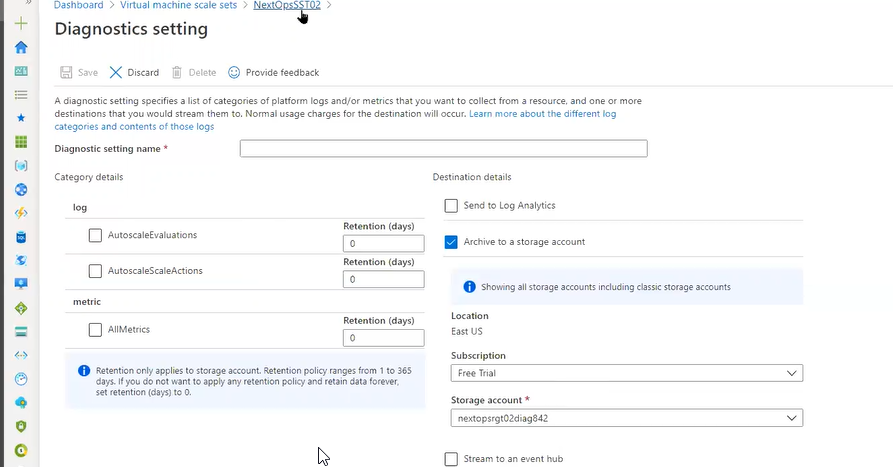




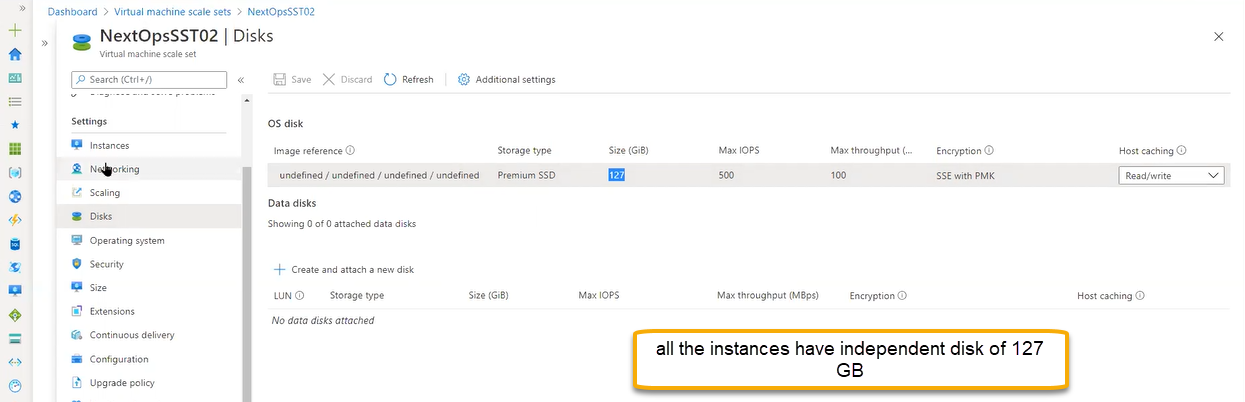


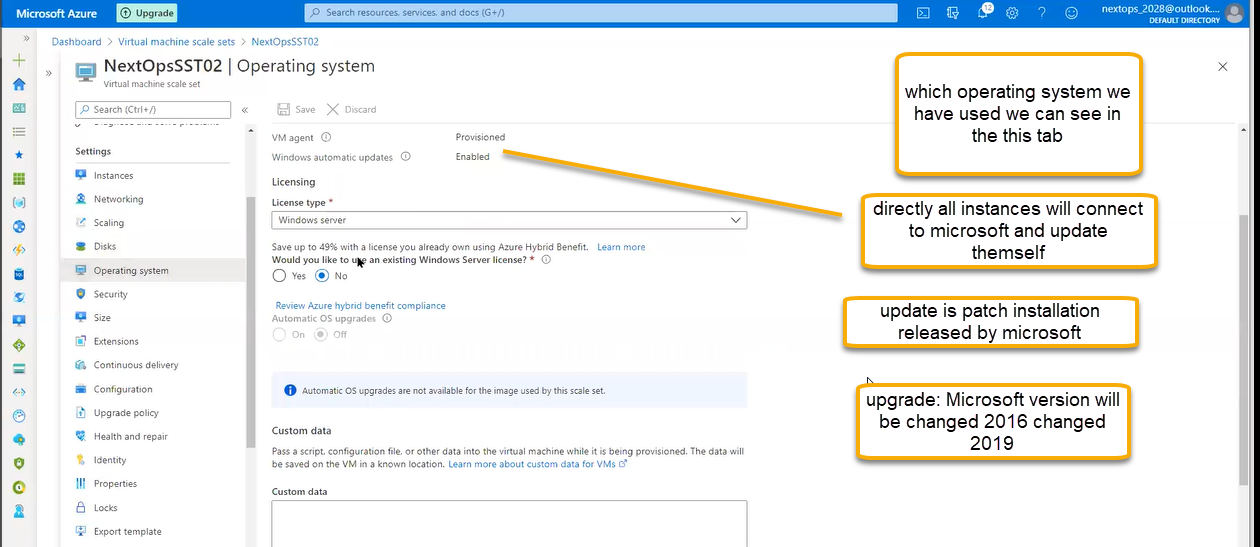


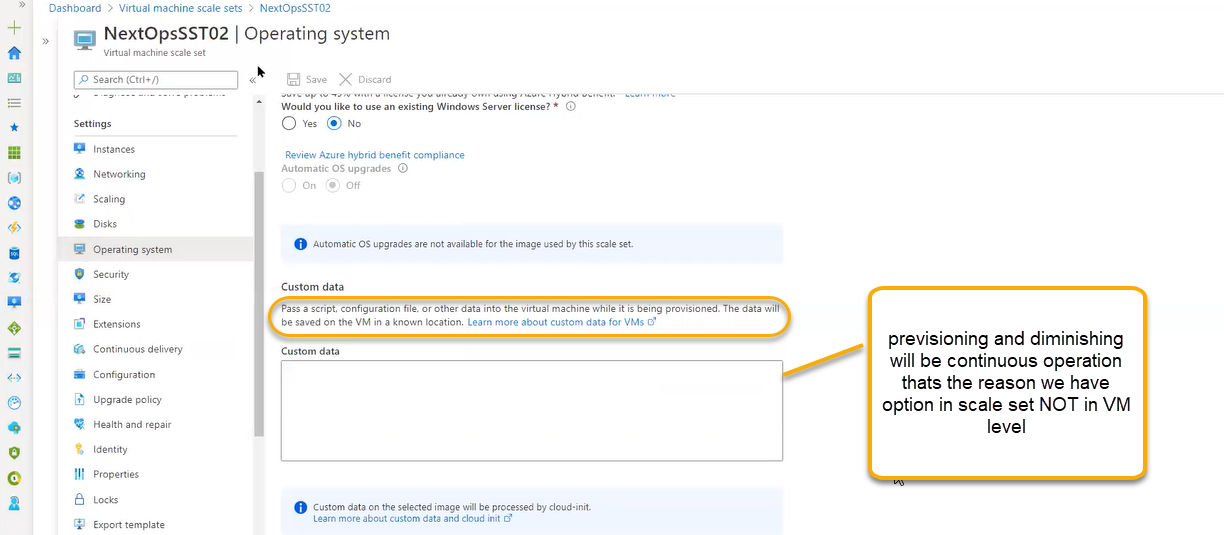




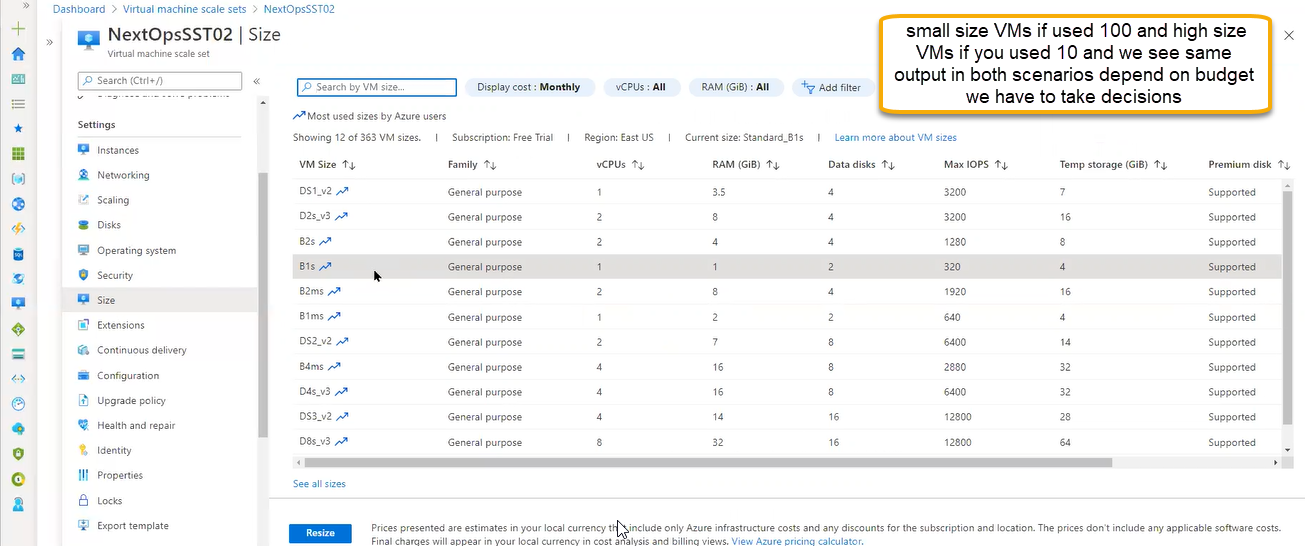
By default it will be saved in activity log and if you want to redirect we can do it here

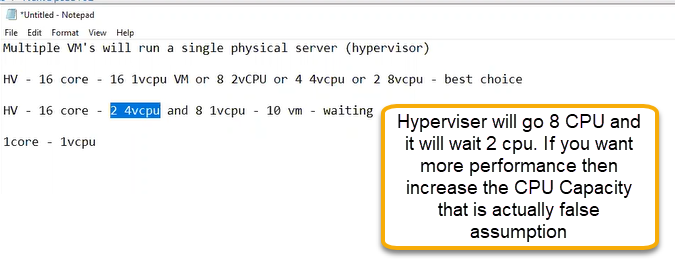






Security will show the recommendations for the scale set





HV-Hypervisor

