

# Stories 10: Virtual Machine Scale Set

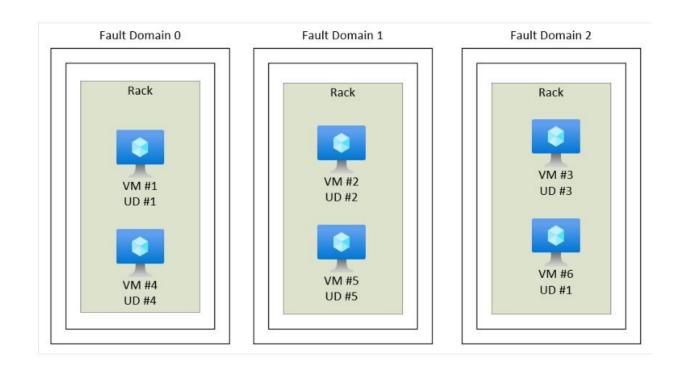
A Virtual Machine Scale Set (VMSS) is a feature in Azure, Microsoft's cloud computing platform, that allows you to deploy and manage a set of identical, load-balanced virtual machines. This is particularly useful for handling high traffic or providing high availability for applications.

**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 the following key benefits:

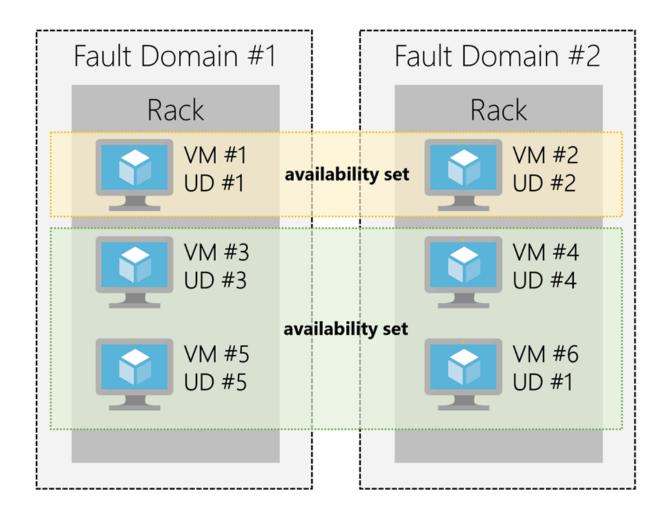
- Easy to create and manage multiple VMs
- Provides high availability and application resiliency by distributing VMs across availability zones or fault domains
- Allows your application to automatically scale as resource demand changes
- Works at large-scale

Fault domain is a concept used to increase the resilience and availability of applications by ensuring that resources are distributed across different physical segments within a data center. Azure uses fault domains as part of its strategy to mitigate the impact of hardware or infrastructure failures.

Stories 10: Virtual Machine Scale Set



### Update domain :



To Create Virtual machine Scale set

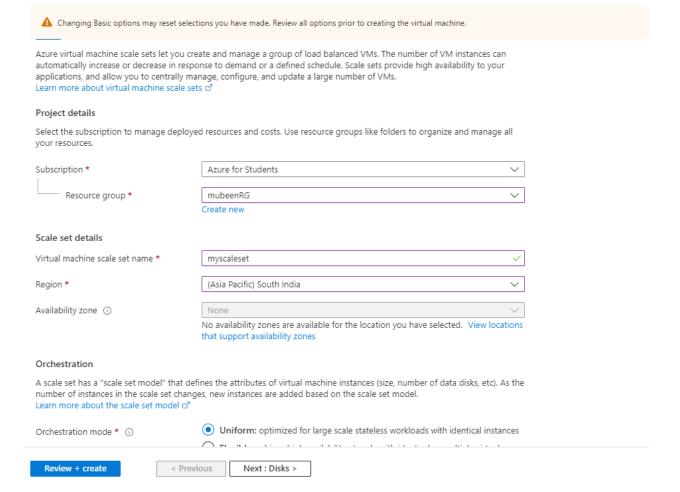
Solution:

# Step 1:

Search for vmss click on Virtual machine scale set

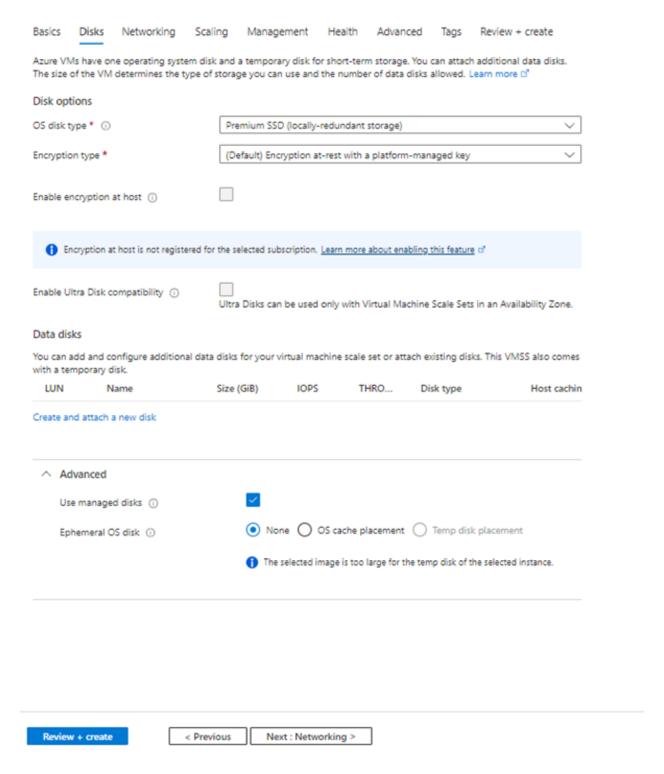
Create

### Create a virtual machine scale set ....



### Adding storage or disks

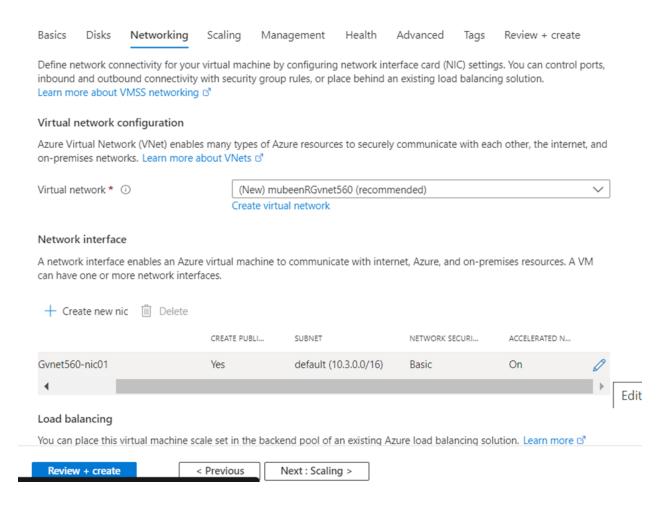
### Create a virtual machine scale set ....



Adding Network to virtual machine scale set

Stories 10: Virtual Machine Scale Set 5

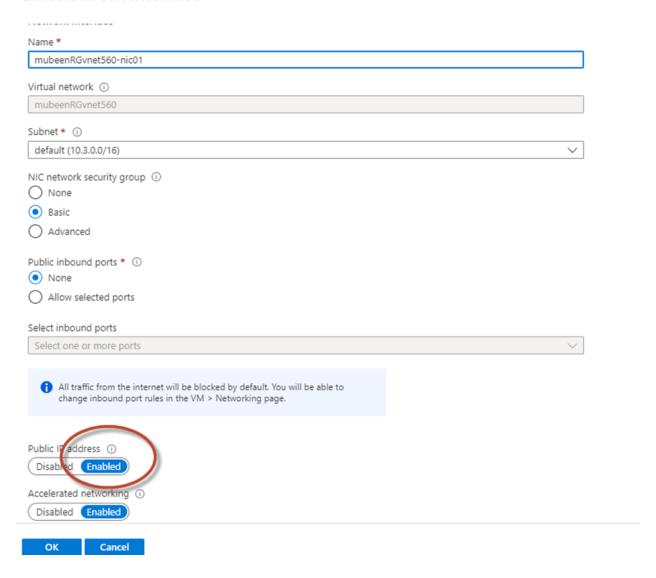
### Create a virtual machine scale set



Note: By default the public Ip address for nic will be disabled so we need to enable it by clicking the pen icon as shown in the above image

Stories 10: Virtual Machine Scale Set

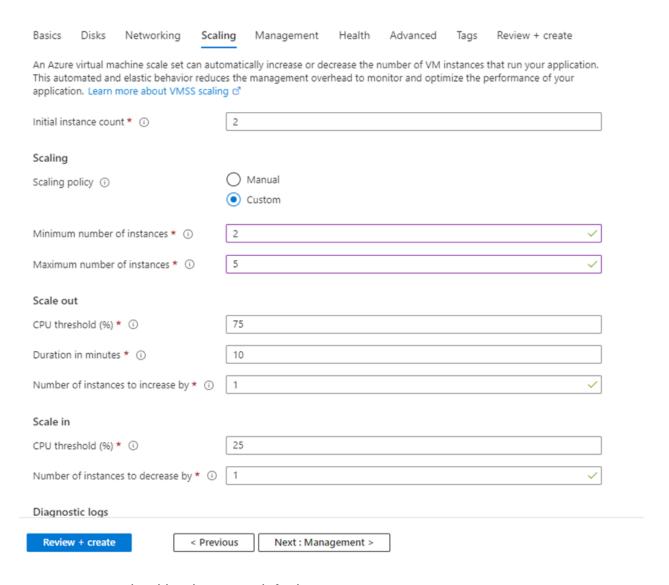
### Edit network interface ....



Configuring the scaling for vmss

Select CUSTOM

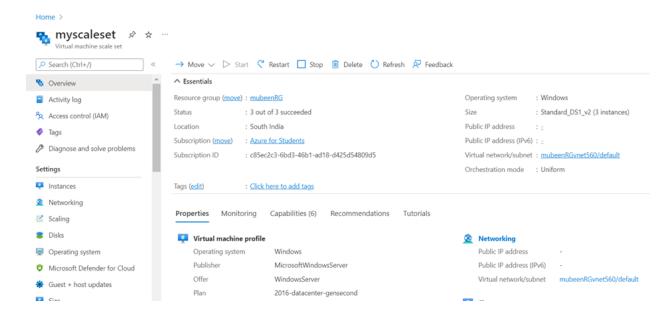
### Create a virtual machine scale set ....



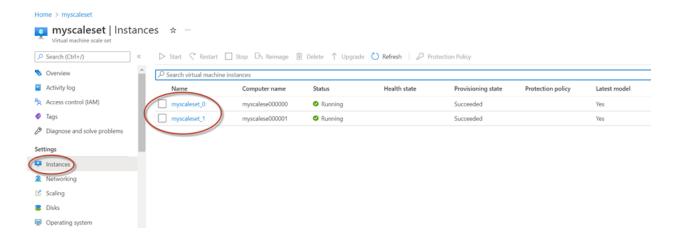
Management > health > leave as default

## Step 2:

After successfully deployed goto



#### Click on instance to list the instance



2 instance are running because we have give minimum 2 instance at scaling option it will scale upto 5 as we have given 5 max scale

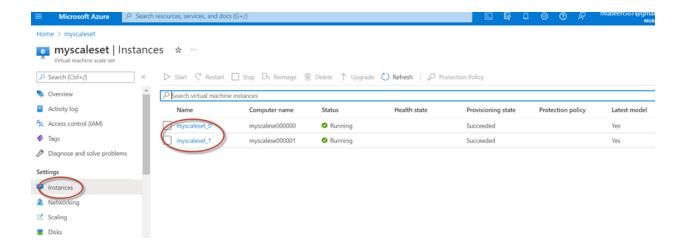
### Step 3:

Connect to one the machine...

Note: Allow 3389 in security Group to connect

To connect and to test the scaling you can RDP it and perform Load testing installing software Heavy load. In one of the VM of scale set

### Click on instances to list the scale set instances



As we have given minimum 2 and max scale 5 when cpu load increases more than 75 percent one more VM will be added to scale set.

stress commands

### Method 1:

### Installation:

If <u>stress</u> is not already installed on your Ubuntu system, you can install it using the package manager. Open a terminal and run:

sudo apt update sudo apt install stress

### **CPU Stress Test:**

To stress the CPU, use the following command:

stress --cpu 4

### Method 2:

top command to check the cpu

Increase cpu load command yes > /dev/null &

Kill command to kill the process

kill -9 <PID>