Azure Section - Virtual machine Scale set.

Scale set is set or a cluster of Virtual Machines which are identical to each other.

All the virtual machines are made from the same image called as base image.

The number of VMs which were added initially, during the time of creation of scale set is called as desired number of VMs.

The scale set first monitors the load of the traffic on this cluster servers and check the aggregate CPU usage of entire VM cluster.

When the aggregated usage goes above the threshold (which is defined at the time of creation of scale set.) then the scale set uses its base image and add extra VMs to the cluster.

This newly added VMs are also identical to the previous VMs which were added at the time of creation of scale set.

And the number of VMs which it adds after the load increases is depend on the predefined value given to the portal at the time of creation of scale set.

Once the base image of scale set is created we should not update it or install any new app to it.

In order to update the VMs in the scale set we first need to create a new base image with all the updates in it and then change old base image inside the scale set with the new one. So that then the VMs which will get created further on will be created from updated image.

Going forward, this cluster of VMs is having an application deployed on it. And every VM is having its own public and private IP address. And when the user is trying to access the web application then how is it possible for the user to figure out to which VM it needs to hit?

Hence here we use LOAD BALANCER.

Load balancer does three things, it checks the incoming traffic and then route it to the VMs inside the cluster depending on the load each VM is handling. The VM with least load will get the current incoming request. Secondly, Load balancer provides single

public IP and host name with which users can access to the website without knowing to which server inside the cluster they are getting hit to!

Finally the load balancer also provides the health check of the VMs in the cluster. And if the VM is not fit then it is terminated and a new VM is added on that place.

We will see this in hands on!