AZURE AVAILABILITY SET

What is an availability set?

Availability sets are logical groupings of VMs that reduce the chance of correlated failures bringing down related VMs at the same time. Availability sets place VMs in different fault domains for better reliability, especially beneficial if a region doesn't support availability zones. When using availability sets, create two or more VMs within an availability set. Using two or more VMs in an availability set helps highly available applications and meets the 99.95% Azure SLA. There's no extra cost for using availability sets, you only pay for each VM instance you create.

Availability sets offer improved VM to VM latencies compared to availability zones, since VMs in an availability set are allocated in closer proximity. Availability sets have fault isolation for many possible failures, minimizing single points of failure, and offering high availability. Availability sets are still susceptible to certain shared infrastructure failures, like datacentre network failures, which can affect multiple fault domains.

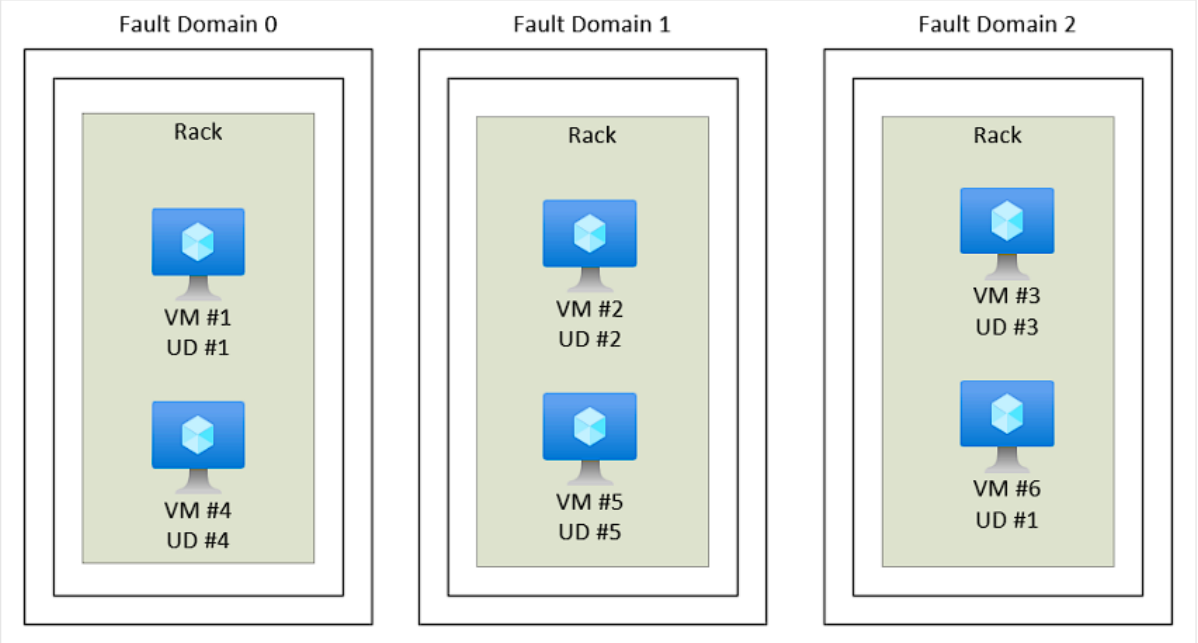
For more reliability than availability sets offer, use availability zones. Availability zones offer the highest reliability since each VM is deployed in multiple datacentres, protecting you from loss of either power, networking, or cooling in an individual datacentre. If your highest priority is the best reliability for your workload, replicate your VMs across multiple availability zones.

How do availability sets work?

Each virtual machine in your availability set is assigned an update domain and a fault domain by the underlying Azure platform. Each availability set can be configured with up to 3 fault domains and 20 update domains. These configurations can't be changed once the availability set has been created. Update domains indicate groups of virtual machines and underlying physical hardware that can be rebooted at the same time. When more than five virtual machines are configured within a single availability set with five update domains, the sixth virtual machine is placed into the same update domain as the first virtual machine, the seventh in the same update domain as the second virtual machine, and so on. The order of update domains being rebooted may not proceed sequentially during planned maintenance, but only one update domain is rebooted at a time. A rebooted update domain is given 30 minutes to recover before maintenance is initiated on a different update domain.

Fault domains define the group of virtual machines that share a common power source and network switch. By default, the virtual machines configured within your availability set are separated across up to three fault domains. While placing your virtual machines into an availability set doesn't protect your application from operating system or application-specific failures, it does limit the impact of potential physical hardware failures, network outages, or power interruptions.

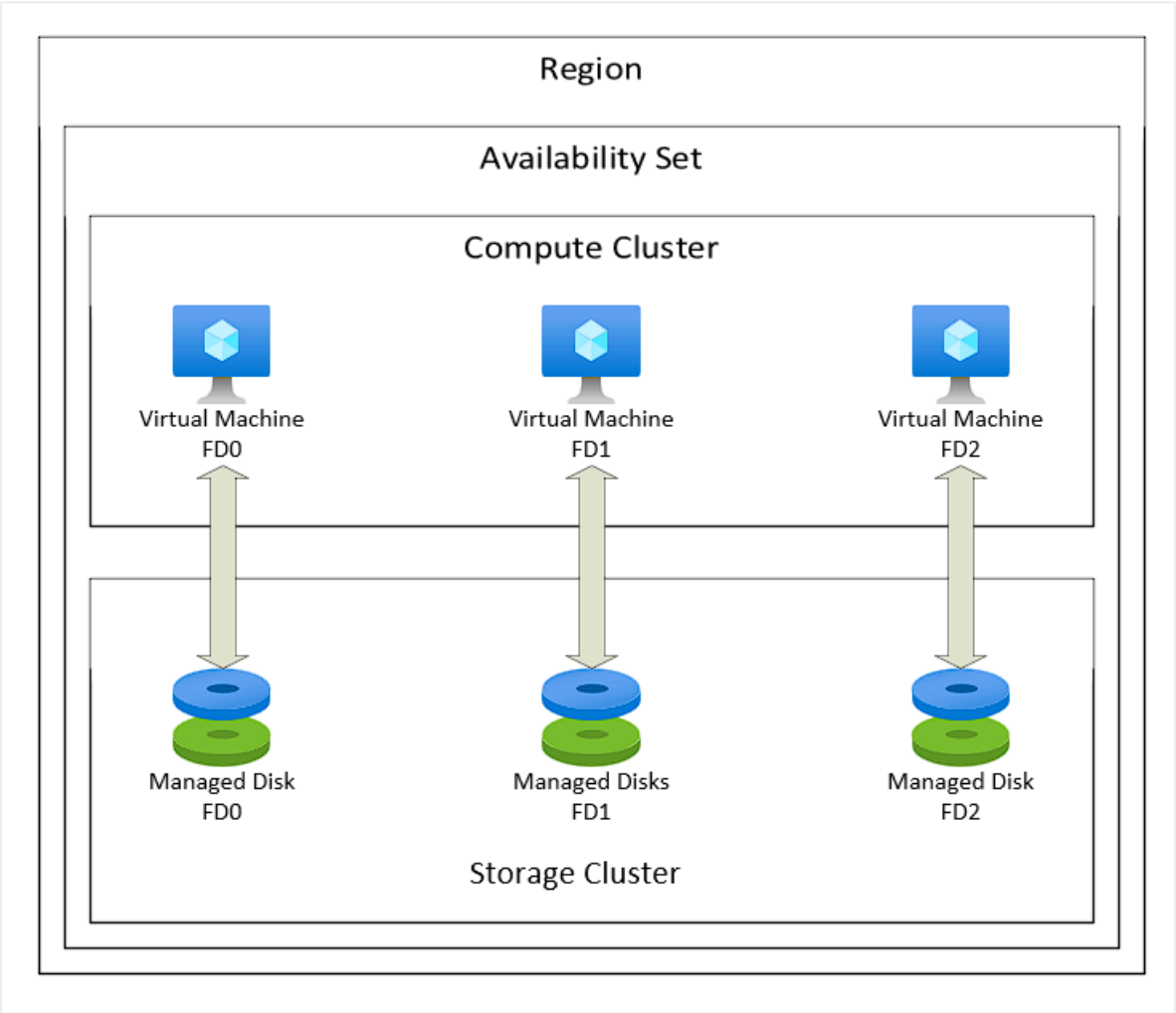
VMs are also aligned with disk fault domains. This alignment ensures that all the managed disks attached to a VM are within the same fault domains.



Under certain circumstances, two VMs in the same availability set might share a fault domain. You can confirm a shared fault domain by going to your availability set and checking the Fault Domain column. A shared fault domain might be caused by the completing following sequence when you deployed the VMs:

1. Deploy the first VM.
2. Stop/deallocate the first VM.
3. Deploy the second VM.

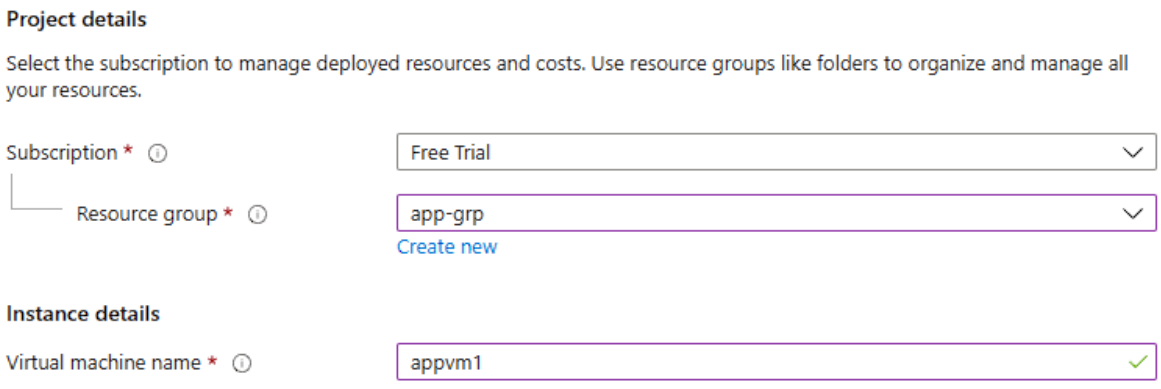
Under these circumstances, the OS disk of the second VM might be created on the same fault domain as the first VM, so the two VMs will be on same fault domain. To avoid this issue, we recommend that you don't stop/deallocate VMs between deployments.



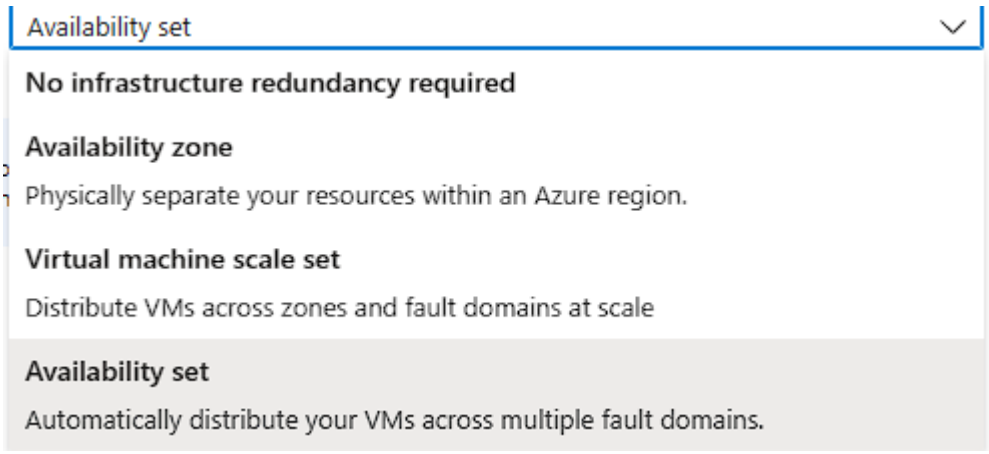
TO BEGIN WITH THE LAB

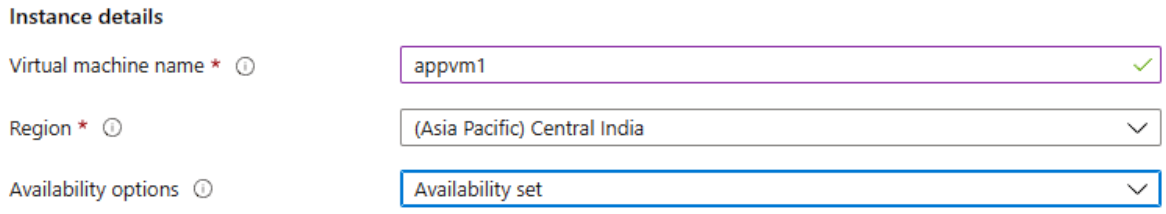
Step 1:

1. Log in to Azure Portal.
2. Go to create resources page.
3. There you need to create Linux Virtual machine.
4. Select your resource group. Then give a name to your virtual machine.

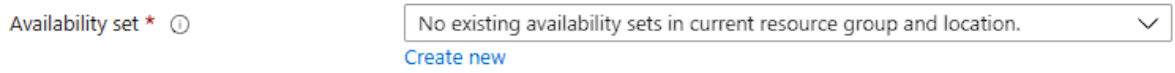


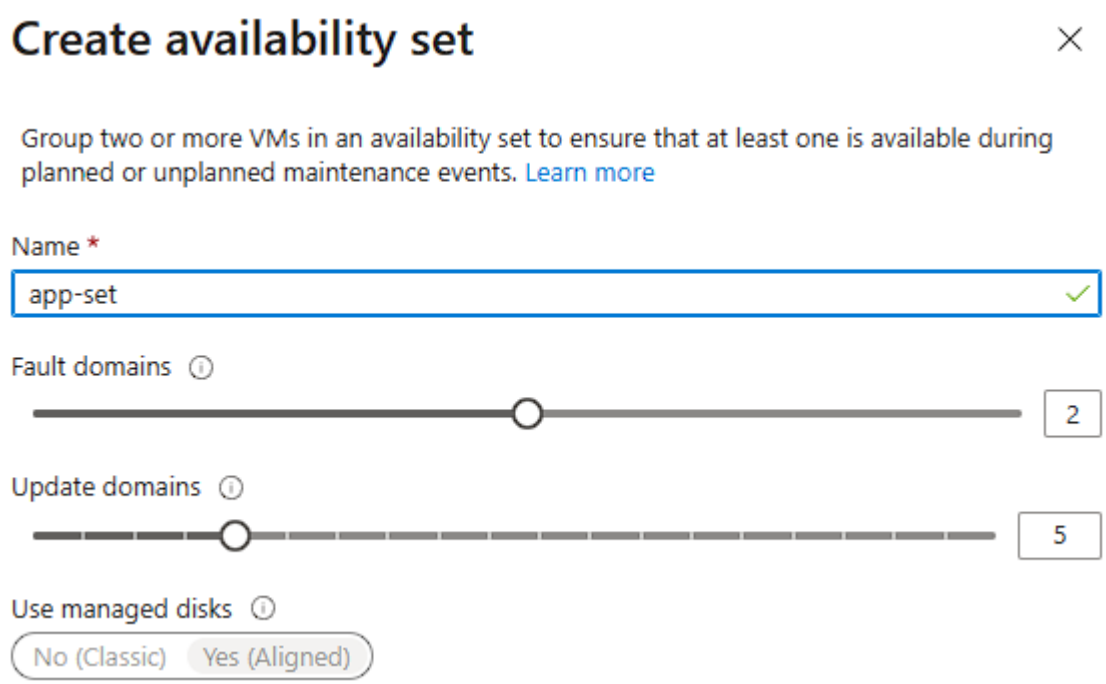
5. Then in the Availability options, you need to select Availability set.



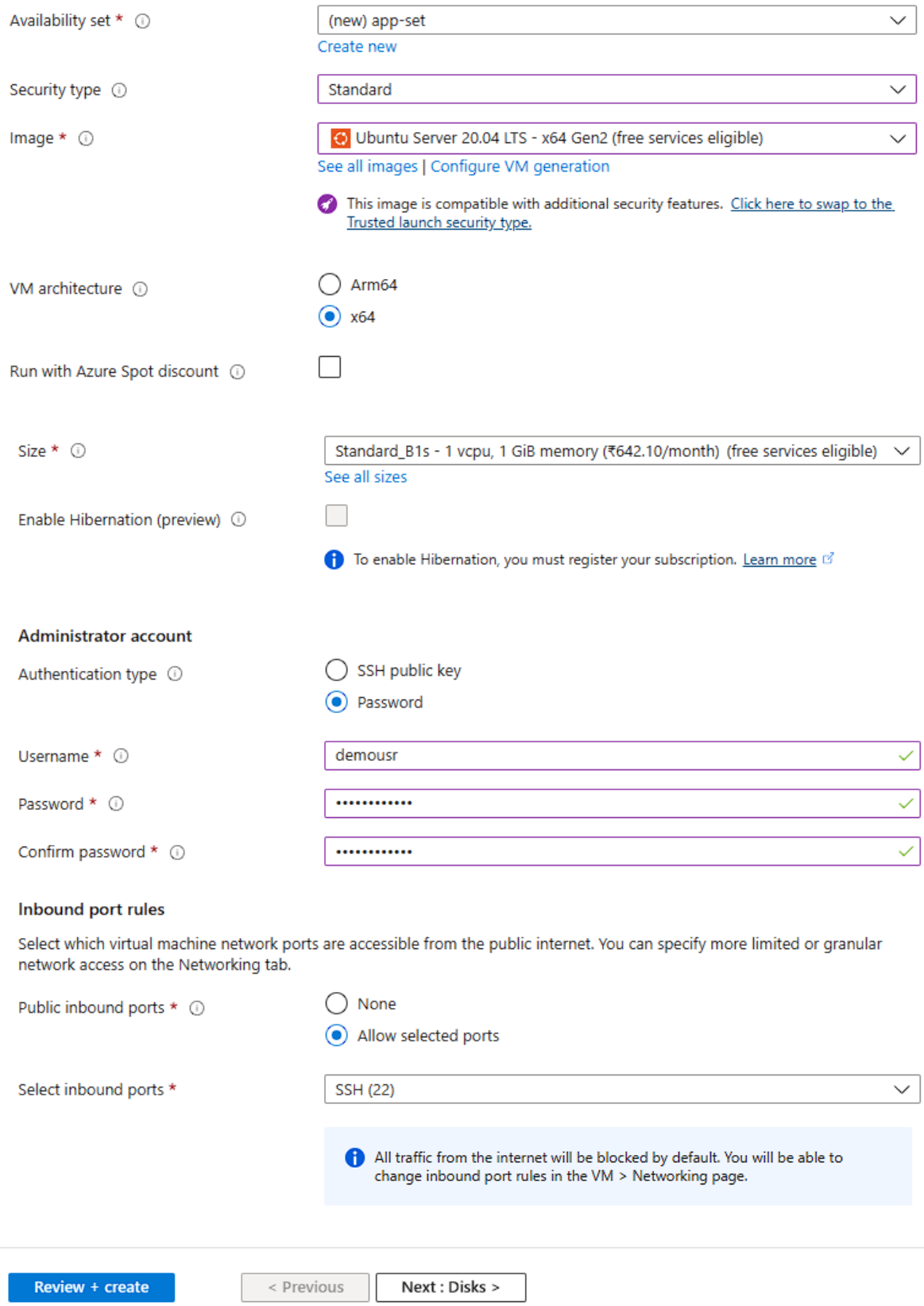


6. Now you need to create an Availability set.
7. For that click on Create new.
8. Give it a name then click on OK (at the bottom)



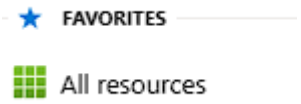


9. Now for the image select Ubuntu server.

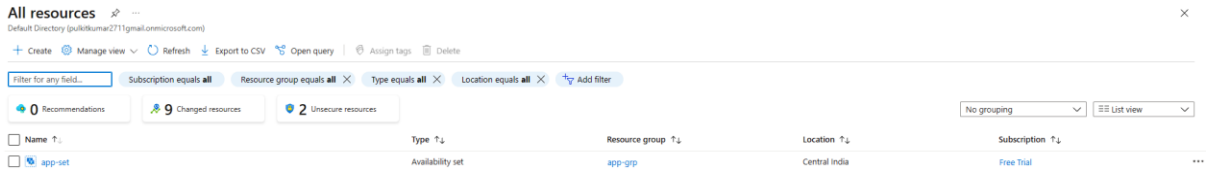


10. Now keep rest of the setting to default and jump to review page and create your machine.

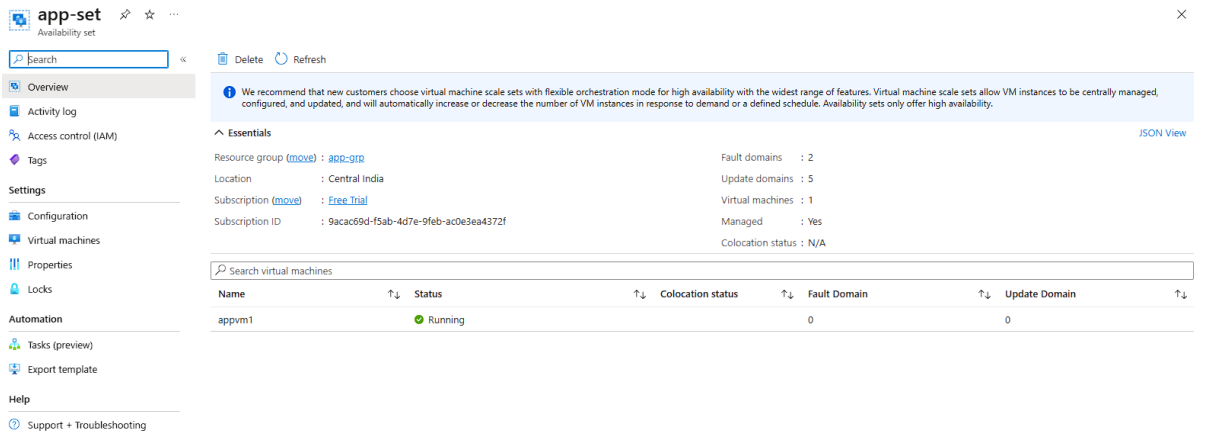
11. Wait for the deployment. Once it gets deployed Go to all resources.



12. In the all resources, you can see your app-set which is your availability set.



13. Open it and there you can see your machine is in running state and it is assigned Fault domain to zero and Update domain to zero.



See, nothing has changed when it comes onto the VM itself, so you can still log onto the VM, you can install your workload, etc. That remains the same. It's only from an infrastructure aspect that now this VM is being made part of something, a logical setting.

Step 2:

1. Now you need to create a new virtual machine with the same **Availability set.**

Create a virtual machine ...

Project details

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

Subscription * : Free Trial

Resource group * : app-grp

Create new

Instance details

Virtual machine name * : appvm2

Region * : (Asia Pacific) Central India

Availability options : Availability set

Based on your input, you might want to consider creating this resource as a virtual machine scale set, which allows you to manage, configure and scale load balanced virtual machines. Create as VMSS

Availability set * : app-set

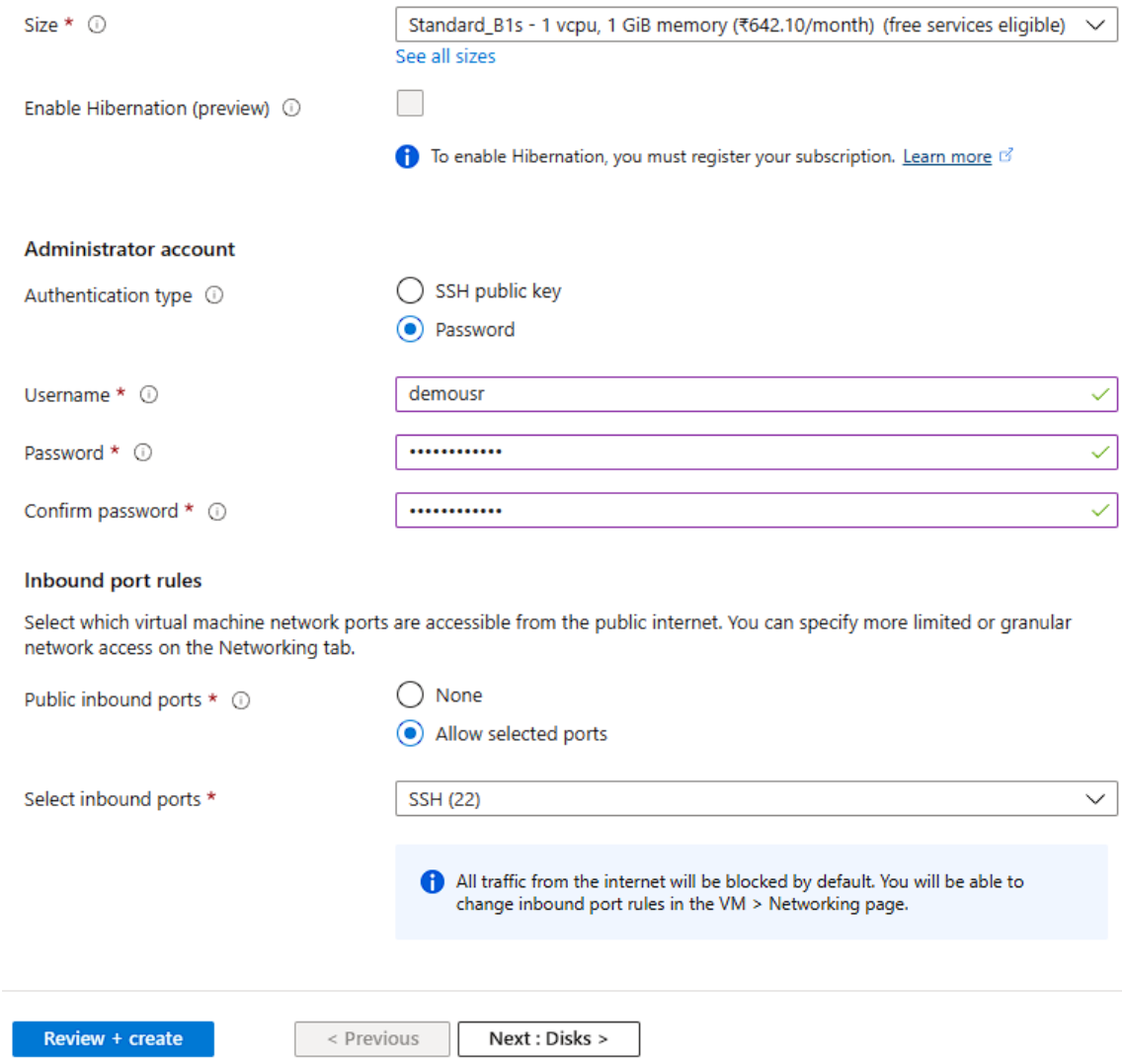
Create new

Security type : Standard

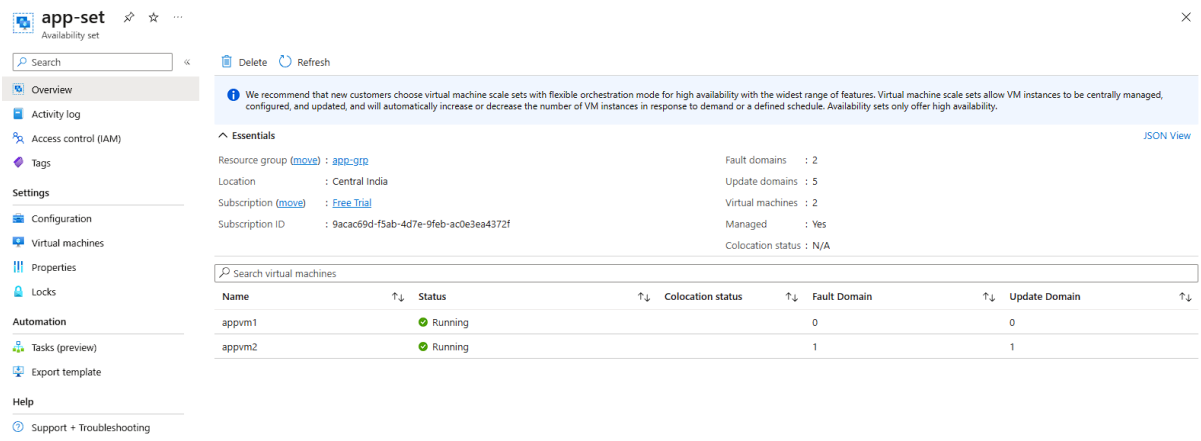
Image * : Ubuntu Server 20.04 LTS - x64 Gen2 (free services eligible)

See all images | Configure VM generation

This image is compatible with additional security features. Click here to swap to the Trusted launch security type.



2. After doing the above setting, jump to review page and create your virtual machine.
3. After the machine has deployed, go to All resources and open app-set which is your availability set.
4. You can see that a new virtual machine can be seen there which is also in the running state.



You will see this VM has been made part of fault domain one and update domain one. So, if any fault occurs on the physical infrastructure that is linked onto, let's say for domain zero, then it would only impact upon one. When Microsoft needs to make updates onto the underlying physical servers, it would first make the update onto, let's say update domain zero. At that point in time, appvm1 may not be available if a restart is required on the underlying physical server. Once that update is complete and appvm1 is in the running state, in case if it was restarted, then the updates would go on to update domain one. And at that point in time, I said if a restart is required, then appvm2 would be restarted accordingly. So again, availability set is just an option to increase the availability of your infrastructure.