Azure Load Balancer

Azure Load Balancer is a Layer-4 network load balancing service that distributes incoming TCP/UDP traffic across multiple backend virtual machines to provide high availability and scalability. It uses frontend IPs, backend pools, health probes, and load-balancing rules to route traffic only to healthy instances.

Main components

→ Frontend IP Configuration

Entry point for traffic.

- Public IP → Internet-facing apps
- Private IP → Internal apps (inside VNet)

→Backend Pool

A group of resources that receive traffic:

- Virtual Machines
- VM Scale Set instances
- NICs / Private IPs

These are your **actual application servers**.

→Health Probes

Continuously checks backend servers (TCP/HTTP):

- If VM is **healthy** → traffic continues
- If VM is **unhealthy** → traffic stops to that VM

This gives **automatic failover**.

→Load Balancing Rules

Define:

- Frontend port (example: 80)
- Backend port (example: 80)
- Protocol (TCP/UDP)
- Which backend pool to use

Types of Load Balancers

→public Load balancer

- Has public ip
- Used for internet-facing apps
- Ex:Web Servers

→Private Load balancere

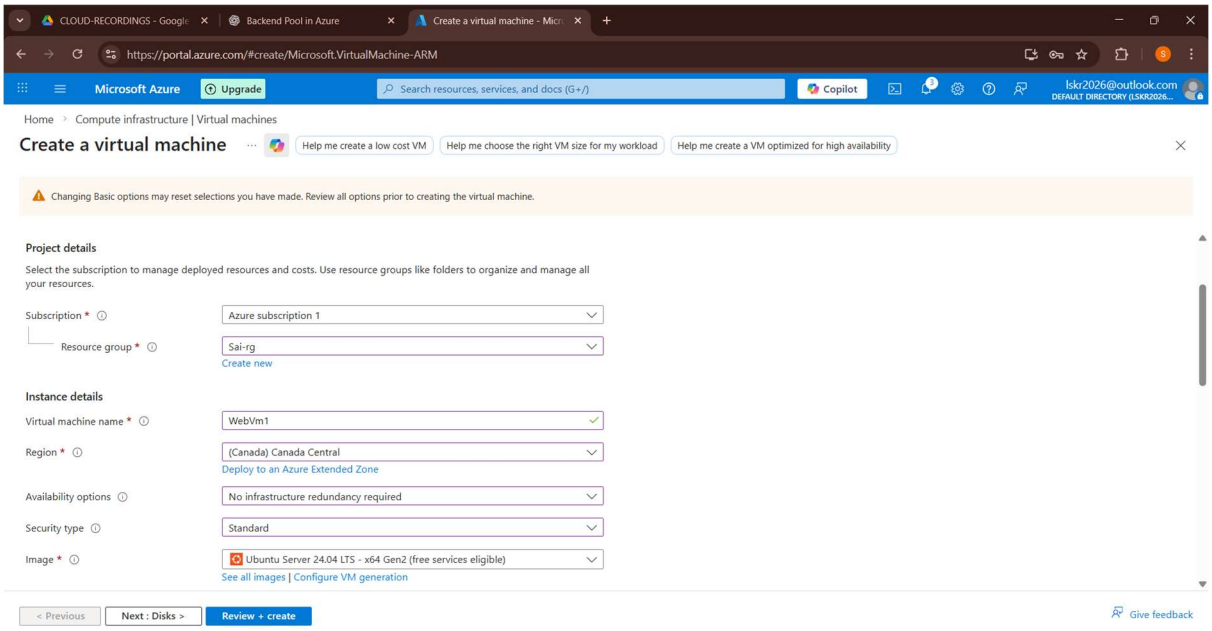
- Uses private IP
- Works only inside Vnet
- Example: app tier → db tier

Steps to create ALB

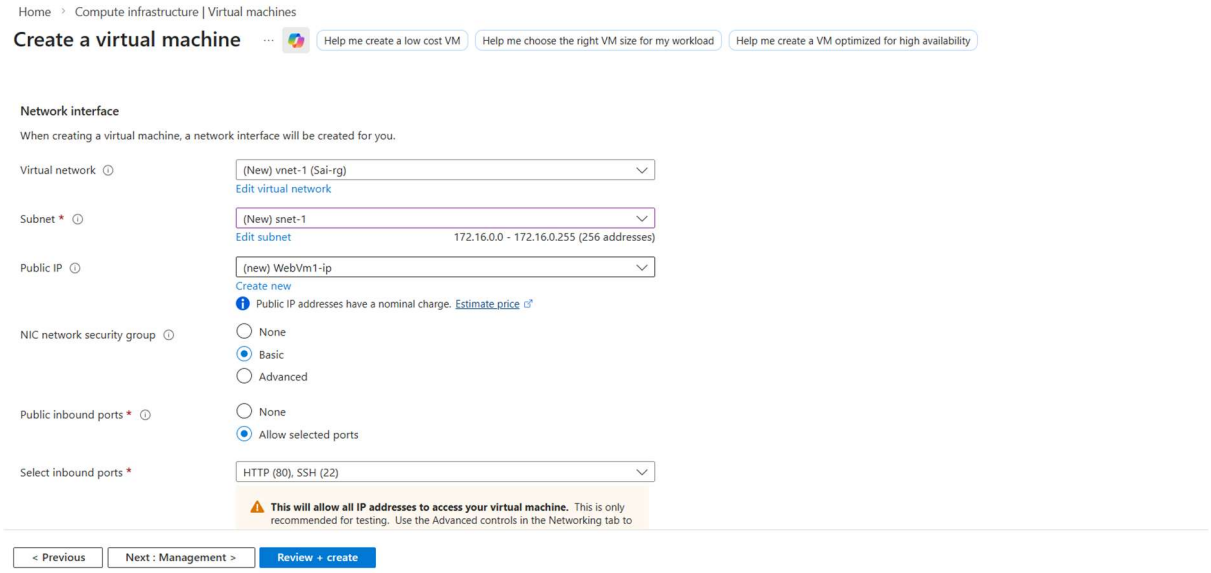
→Lets Create Azure Load Balancer

→Create two Vms with in a vnet with different Subnets

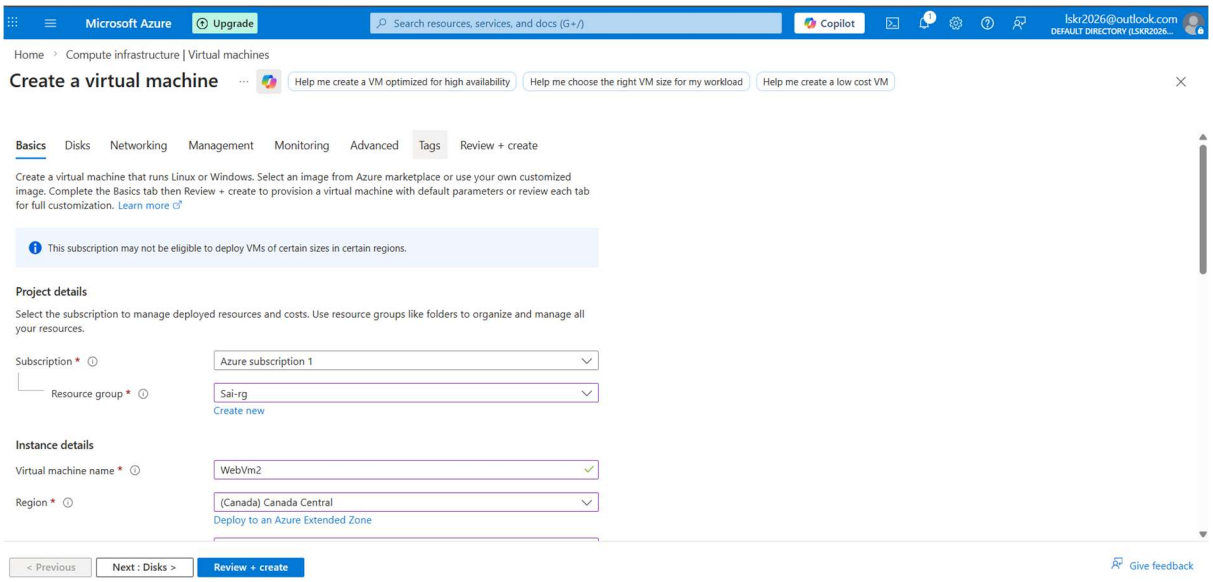
→Creating a WebVm1



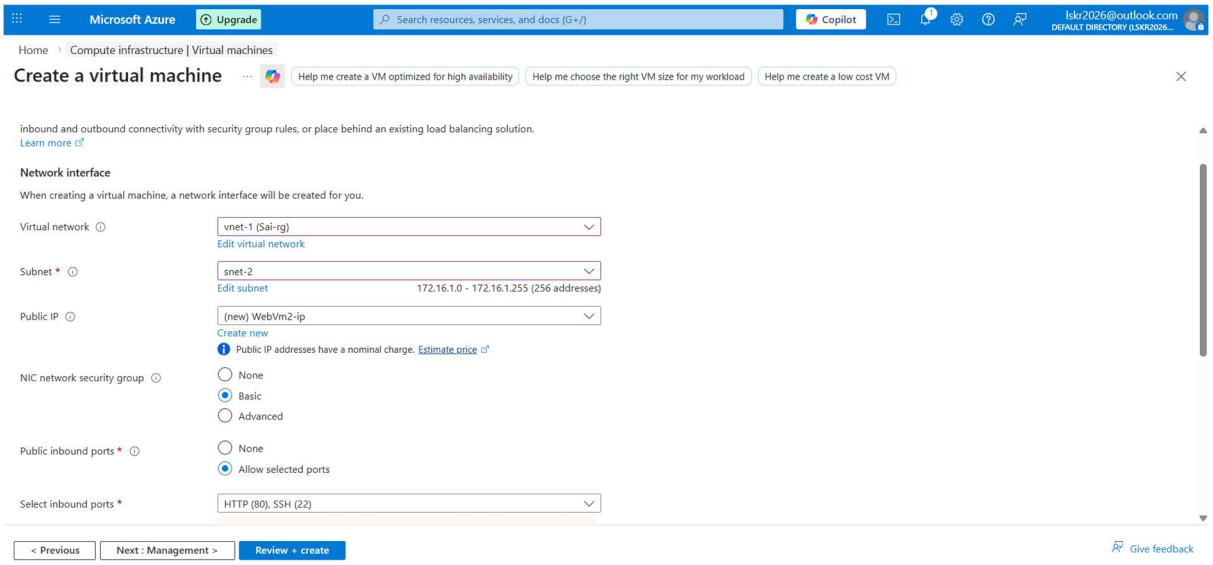
→Adding Vnet and Subnet for WebVm1



→Creating Second WebVm as WebVm2

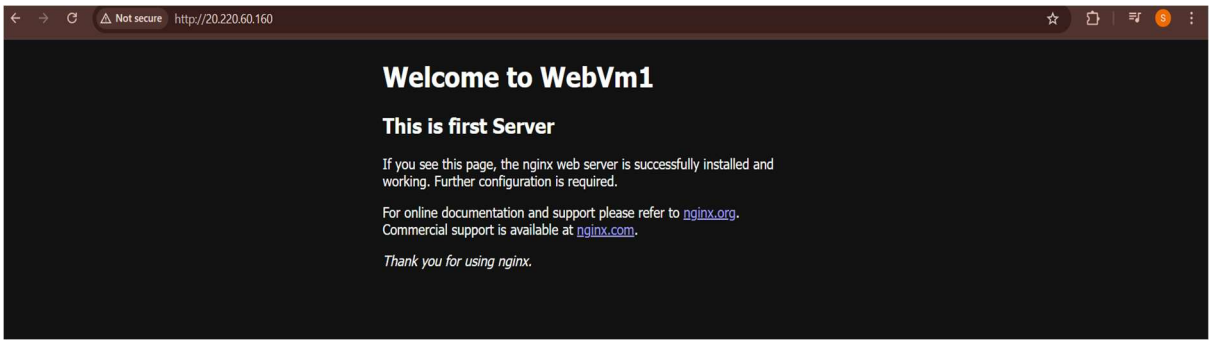


→Adding Vnet and Subnet for WebVm2

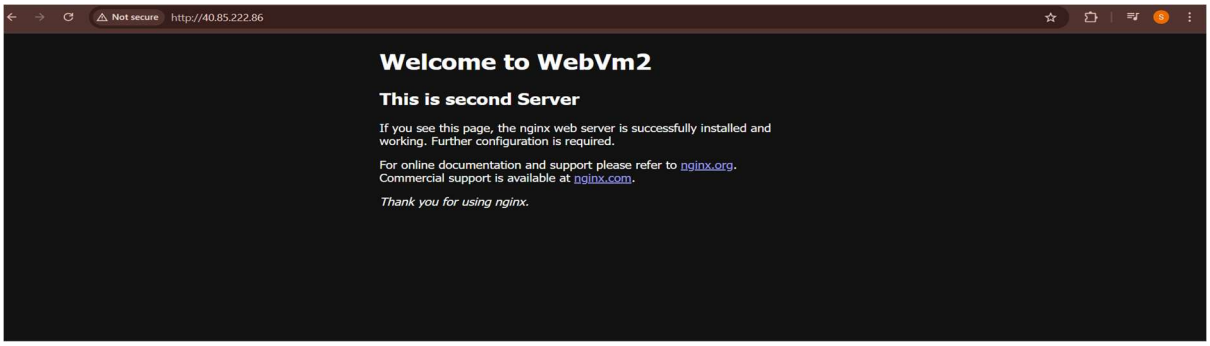


→Then install nginx on two Vms and check it is working or not

→WebVm1

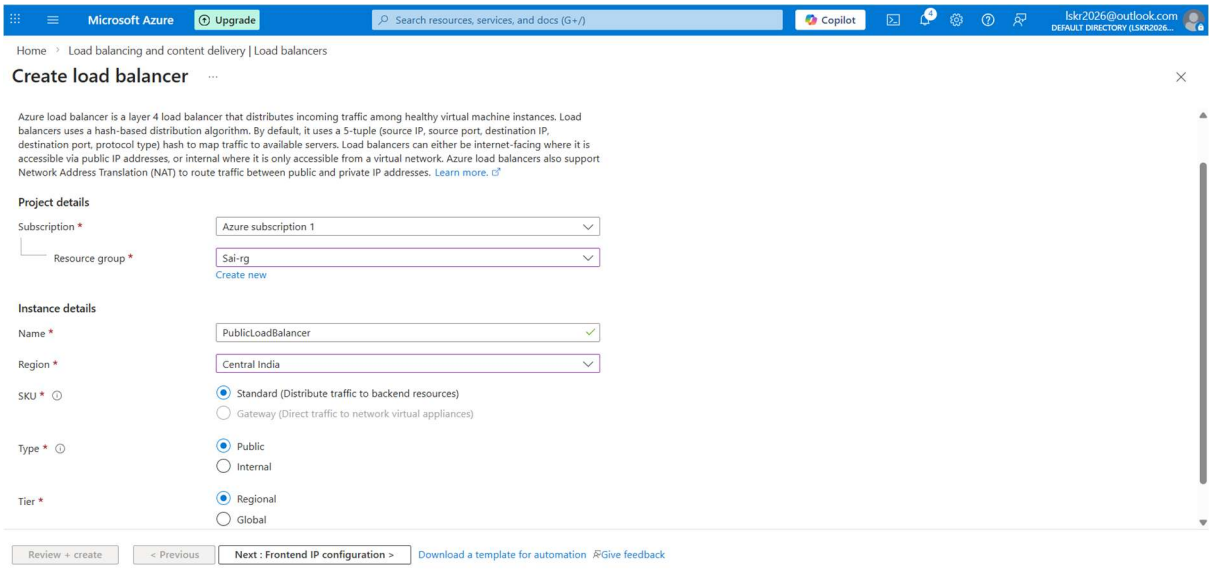


→WebVm2

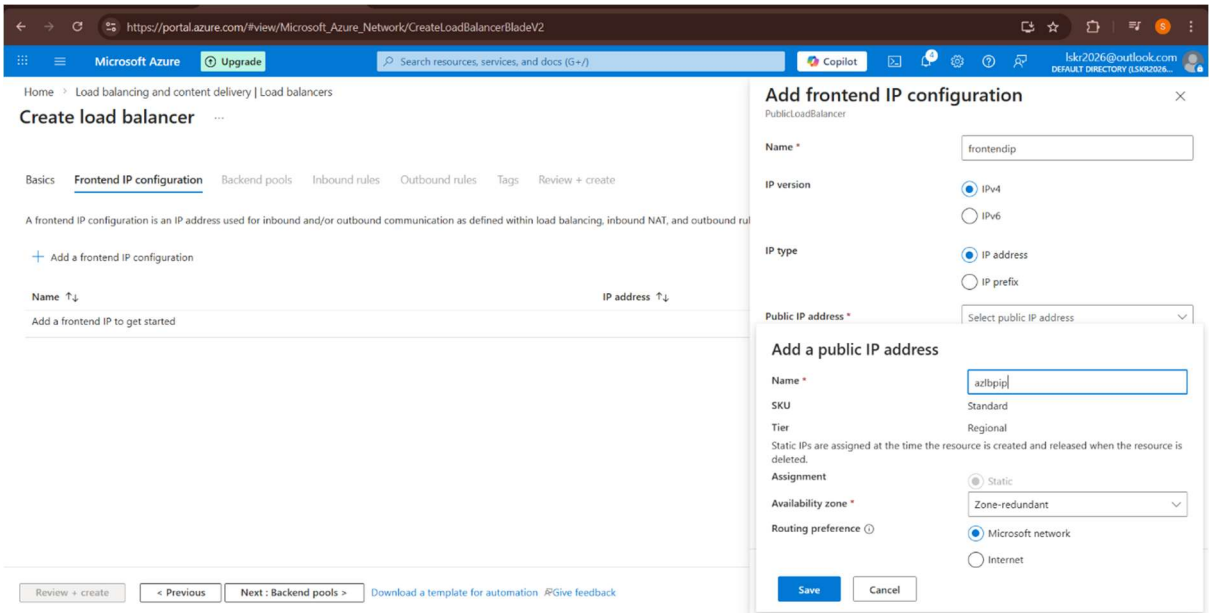


→Search for load balancet and click on create and select standard load balancer

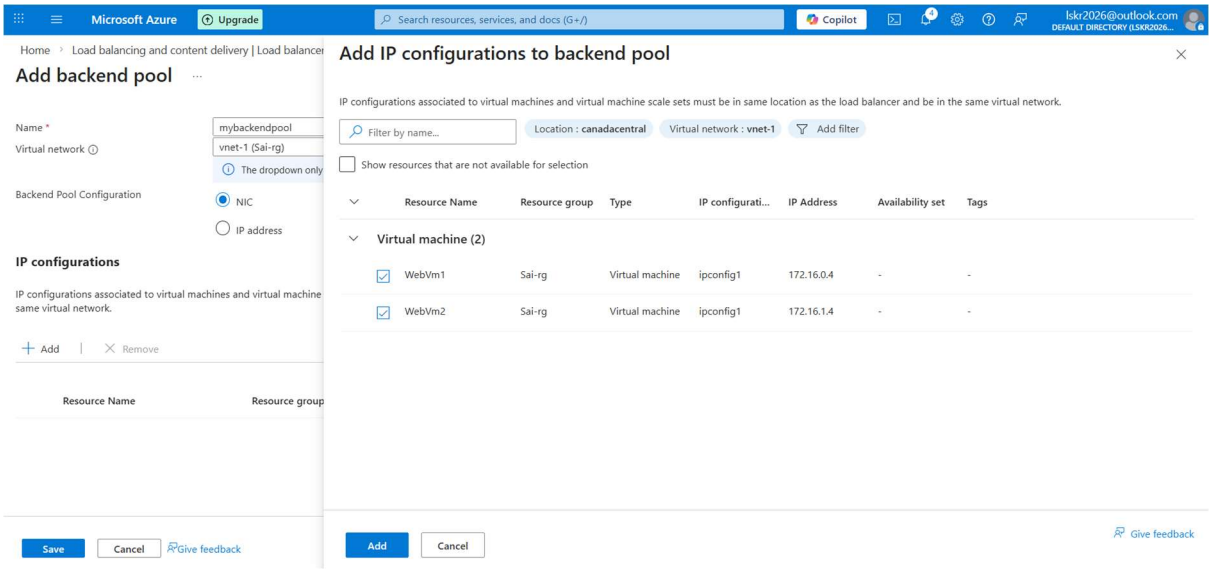
→Creating a Load balancer



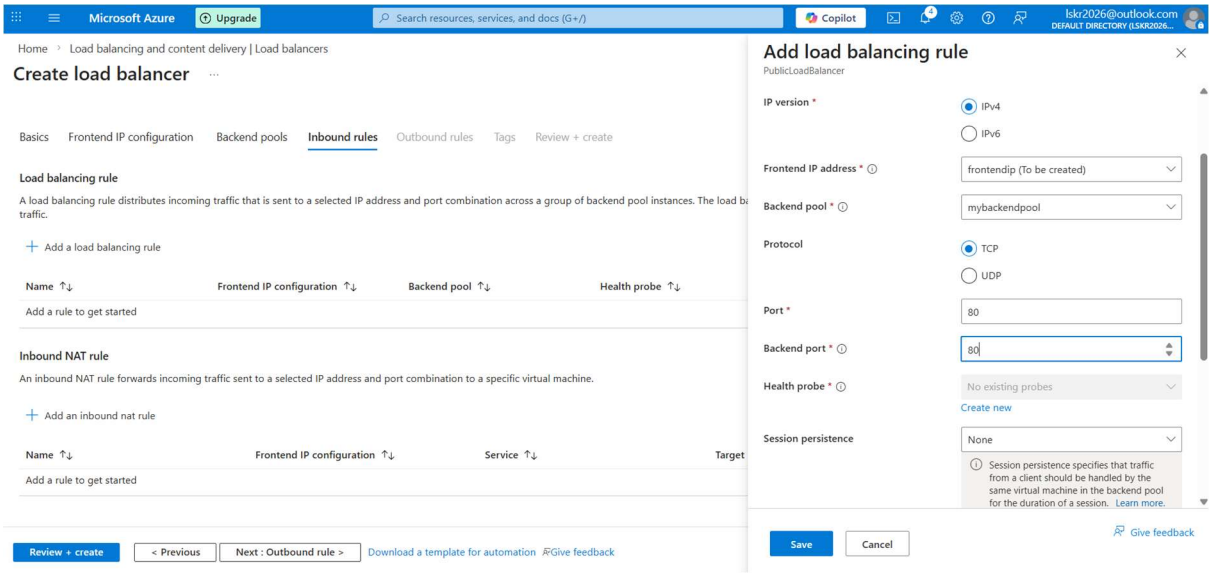
→Adding Frontend ip to the load balancer



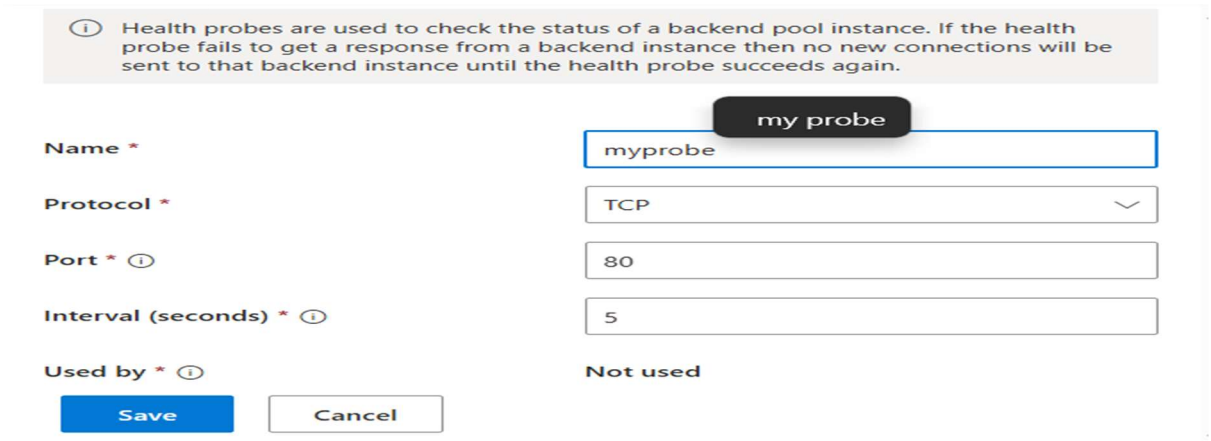
→Adding a backend pool to the load balancer and the two vms for this load balancer



→Add Load Balancing rule



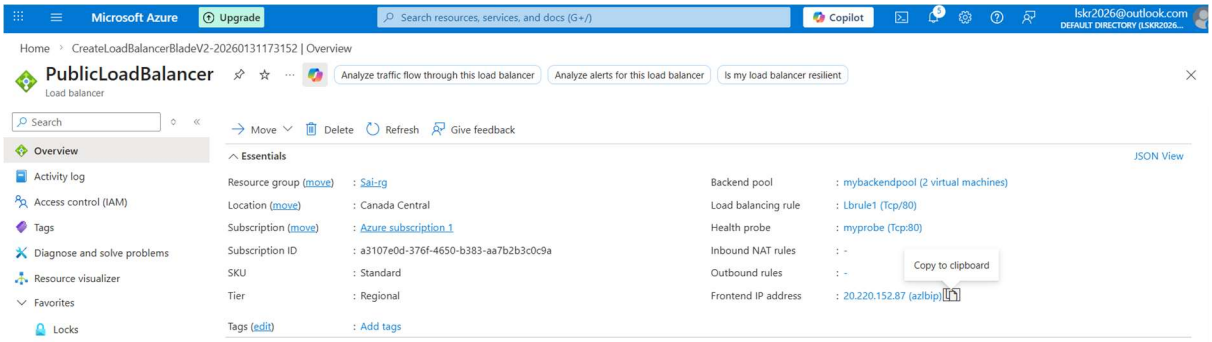
→Adding Health probe



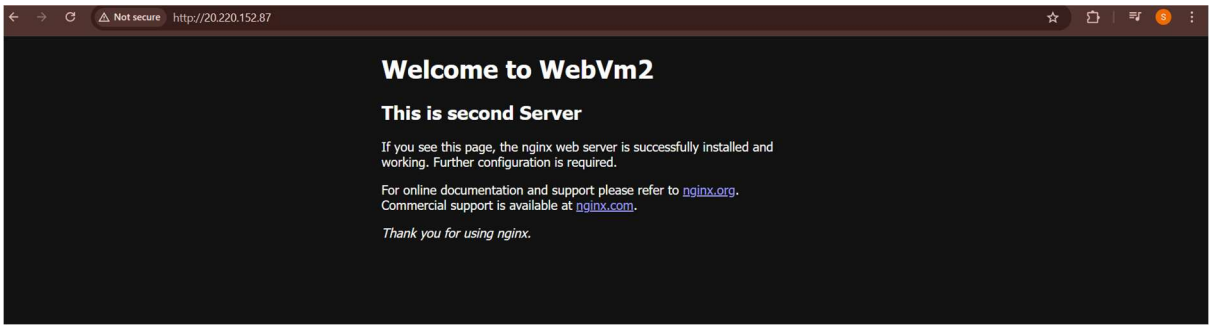
→Check the review and click on create

→The frontend ip is the public ip of your load balancer

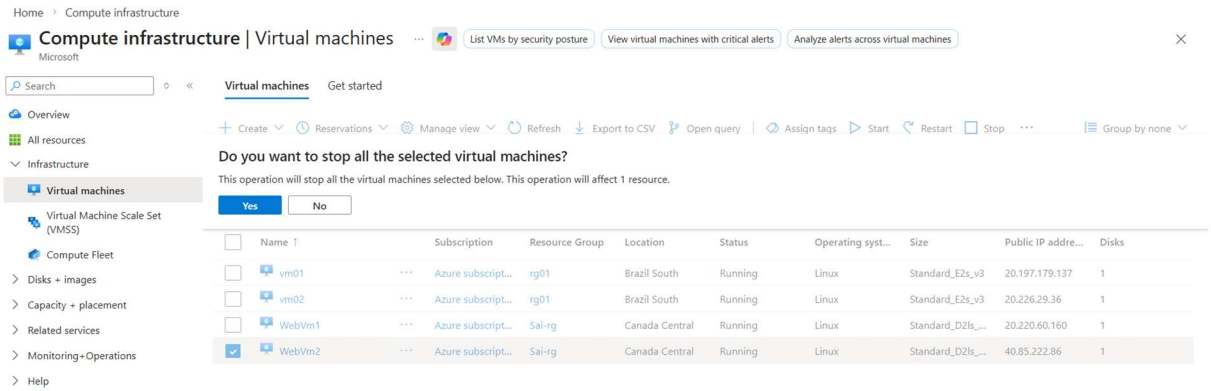
→Copy and search in browser



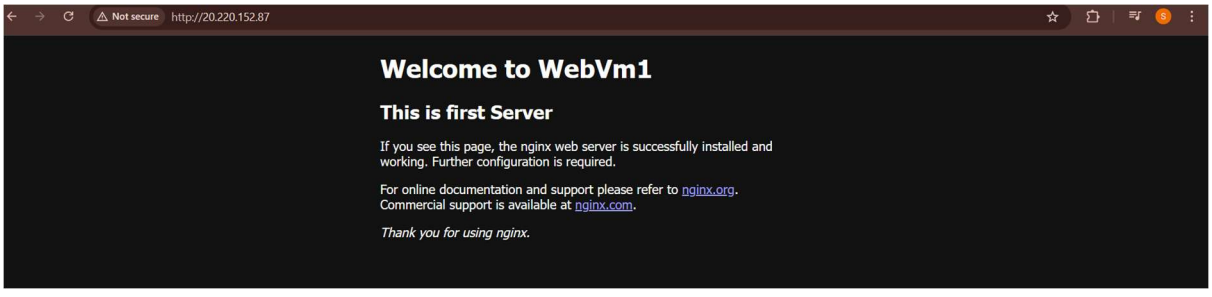
→First I search in this browser the traffic route to my webVm2



→When I stop my WebVm2



→Then traffic routes to my WebVm1



→When you start WebVm2 and search in another browser you can see the webvm2 after some time.