# AKS Egress Traffic

Houssem Dellai

# AKS OutboundType for Egress

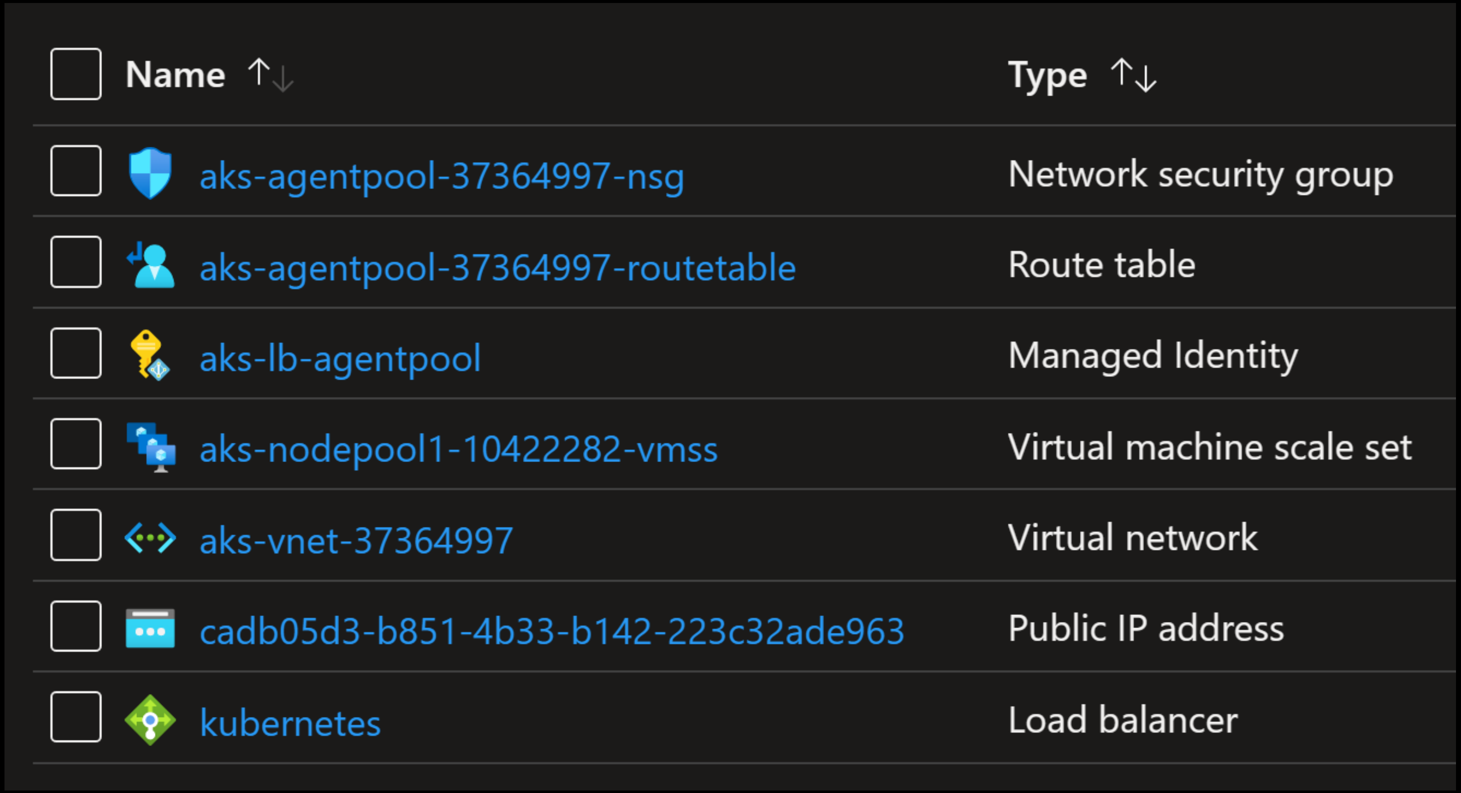
- LoadBalancer
- NAT Gateway
    - ManagedNatGateway
    - UserAssignedNatGateway
- UserDefinedRouting (UDR mode)

# AKS OutboundType LoadBalancer (default)

The load balancer is used for egress through an AKS-assigned public IP.

One or more other public IPs could be used for services with type LoadBalancer.

```
az aks create -g $AKS_RG `
   -n $AKS_NAME `
   --enable-managed-identity `
   --outbound-type loadBalancer
```

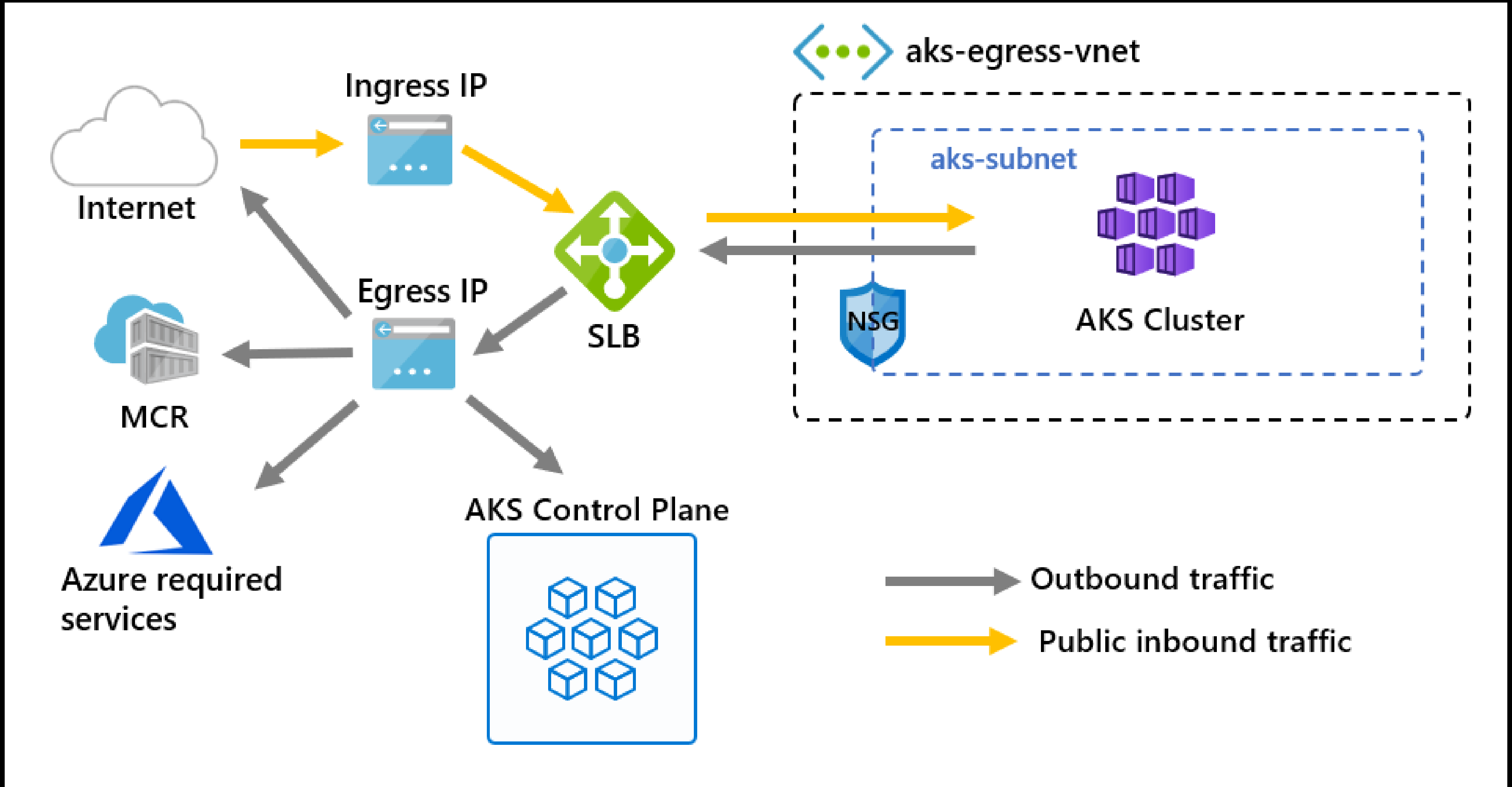


| □ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| □ | aks-agentpool-37364997-nsg | Network security group |
| □ | aks-agentpool-37364997-routetable | Route table |
| □ | aks-lb-agentpool | Managed Identity |
| □ | aks-nodepool1-10422282-vmss | Virtual machine scale set |
| □ | aks-vnet-37364997 | Virtual network |
| □ | cadb05d3-b851-4b33-b142-223c32ade963 | Public IP address |
| □ | kubernetes | Load balancer |

# AKS OutboundType LoadBalancer

One public IP used for egress traffic.

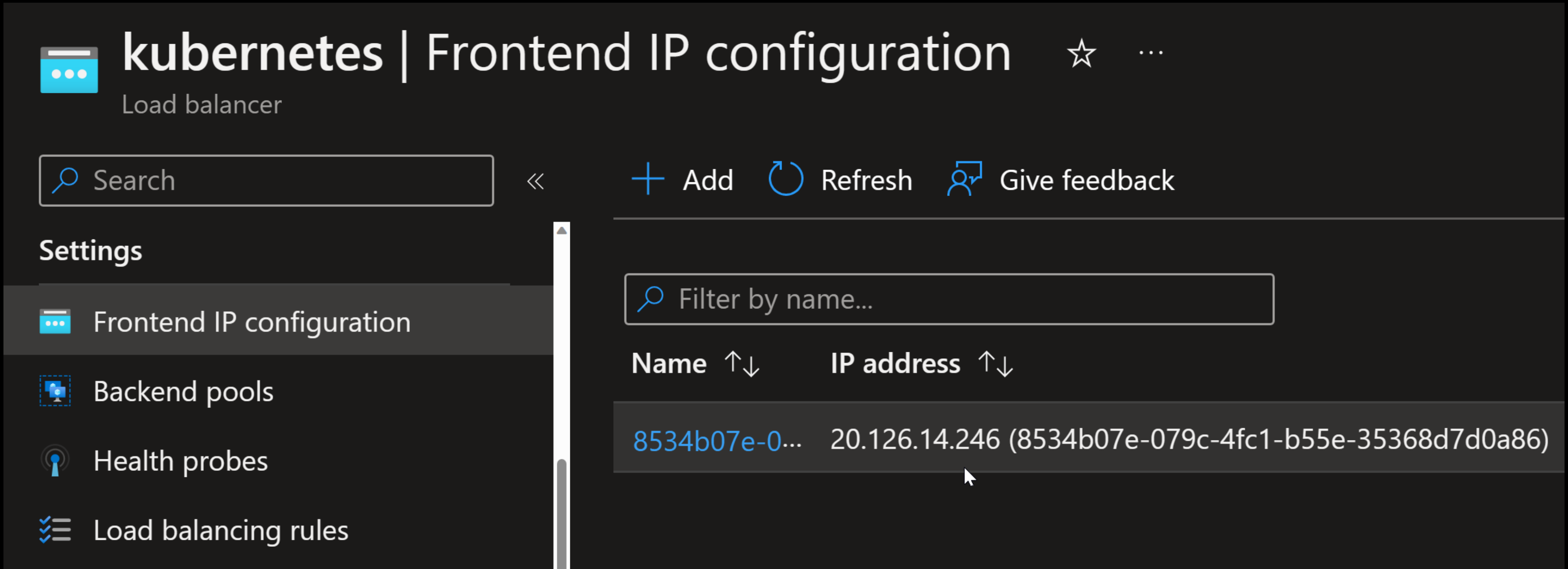
One or more IPs are used for ingress (public services)



# Pods egress through Load Balancer public IP

```
kubectl run nginx --image=nginx
pod/nginx created

kubectl exec nginx -it -- /bin/bash
root@nginx:/# curl ifconfig.me
20.126.14.246
```



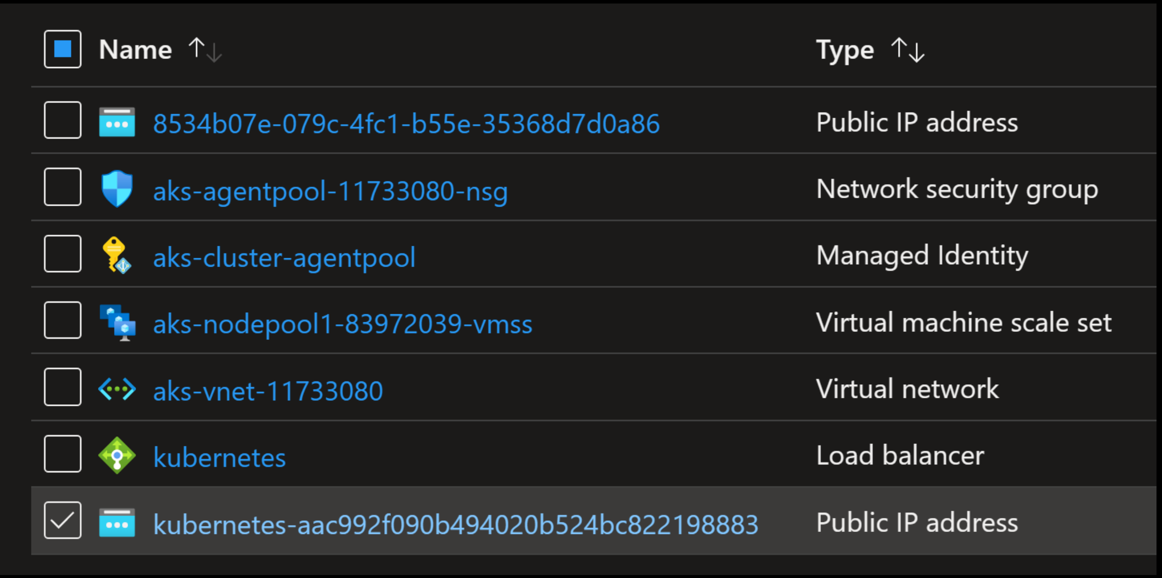
kubernetes | Frontend IP configuration

Load balancer

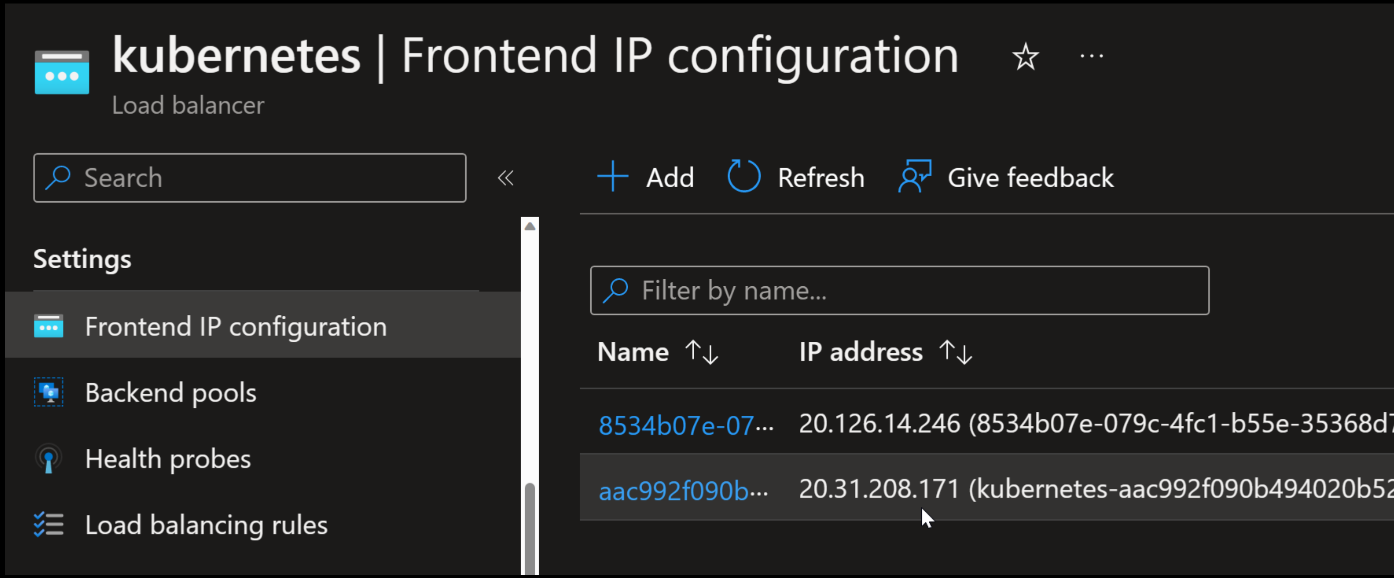
| Name | IP address |
|---|---|
| 8534b07e-0... | 20.126.14.246 (8534b07e-079c-4fc1-b55e-35368d7d0a86) |

# Creating public service creates new public IP in LB

```
kubectl expose deployment nginx --name nginx --port=80 --type LoadBalancer
kubectl get svc
NAME         TYPE           CLUSTER-IP     EXTERNAL-IP     PORT(S)        AGE
kubernetes   ClusterIP      10.0.0.1       <none>          443/TCP        10h
nginx        LoadBalancer   10.0.106.59    20.31.208.171   80:31371/TCP   9s
```



| ■ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| □ | 8534b07e-079c-4fc1-b55e-35368d7d0a86 | Public IP address |
| □ | aks-agentpool-11733080-nsg | Network security group |
| □ | aks-cluster-agentpool | Managed Identity |
| □ | aks-nodepool1-83972039-vmss | Virtual machine scale set |
| □ | aks-vnet-11733080 | Virtual network |
| □ | kubernetes | Load balancer |
| ☑ | kubernetes-aac992f090b494020b524bc822198883 | Public IP address |



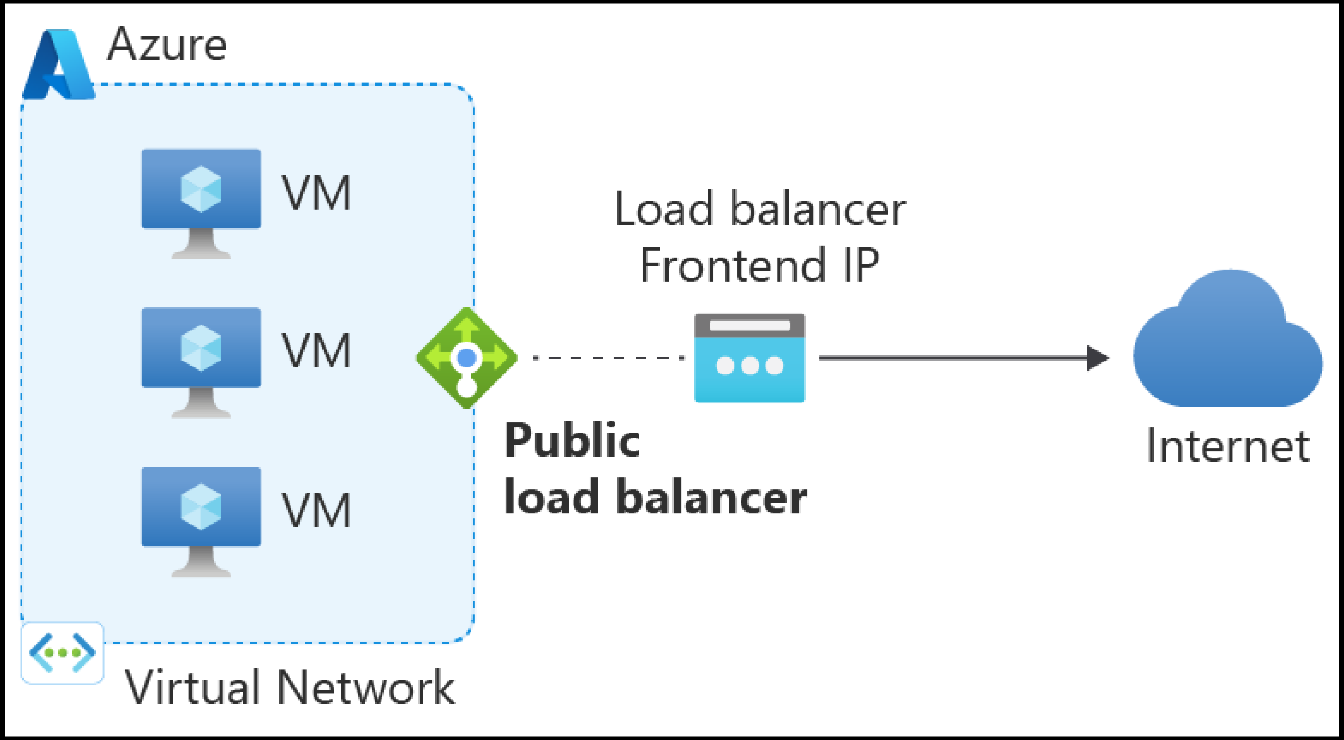
## kubernetes | Frontend IP configuration

Load balancer

| Name ↑↓ | IP address ↑↓ |
|---|---|
| 8534b07e-07... | 20.126.14.246 (8534b07e-079c-4fc1-b55e-35368d7 |
| aac992f090b... | 20.31.208.171 (kubernetes-aac992f090b494020b52 |

# Load Balancer SNAT port exhaustion issue

The frontend IPs of a public load balancer can be used to provide outbound connectivity to the internet for backend instances. This configuration uses source network address translation (SNAT) to translate virtual machine's private IP into the load balancer's public IP address.



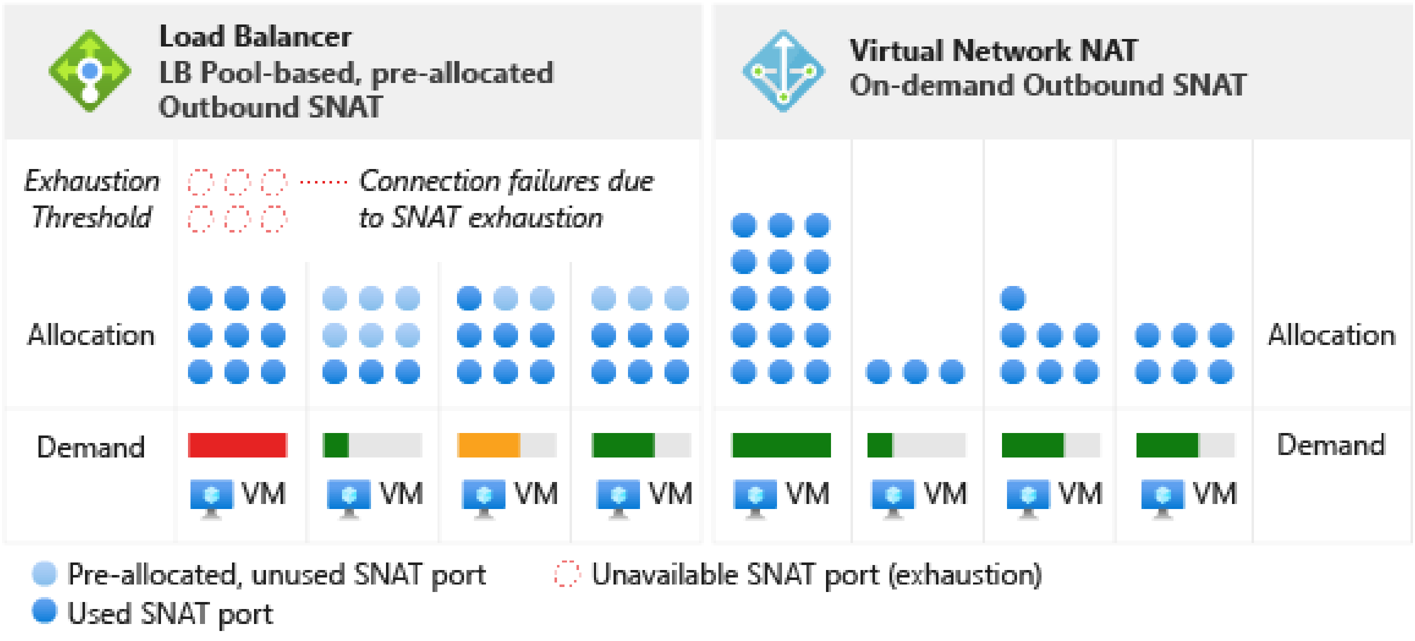
# Load Balancer SNAT port exhaustion issue

With LB, each VM use a fixed number (up to 1024) pre-allocated SNAT ports.

If a VM need more, it will run into port exhaustion and connection will be dropped.

Meanwhile, other VMs might have available SNAT ports!

With NAT Gateway, pre-allocation of SNAT ports isn't required, which means SNAT ports aren't left unused by VMs not actively needing them.


Load Balancer
LB Pool-based, pre-allocated Outbound SNAT
Virtual Network NAT
On-demand Outbound SNAT
Exhaustion Threshold
Connection failures due to SNAT exhaustion
Allocation
Demand
VM
VM
VM
VM
VM
VM
VM
VM
Allocation
Demand
Pre-allocated, unused SNAT port
Used SNAT port
Unavailable SNAT port (exhaustion)

https://learn.microsoft.com/en-us/azure/virtual-network/nat-gateway/nat-gateway-resource

# Load Balancer SNAT port exhaustion (solution 1)

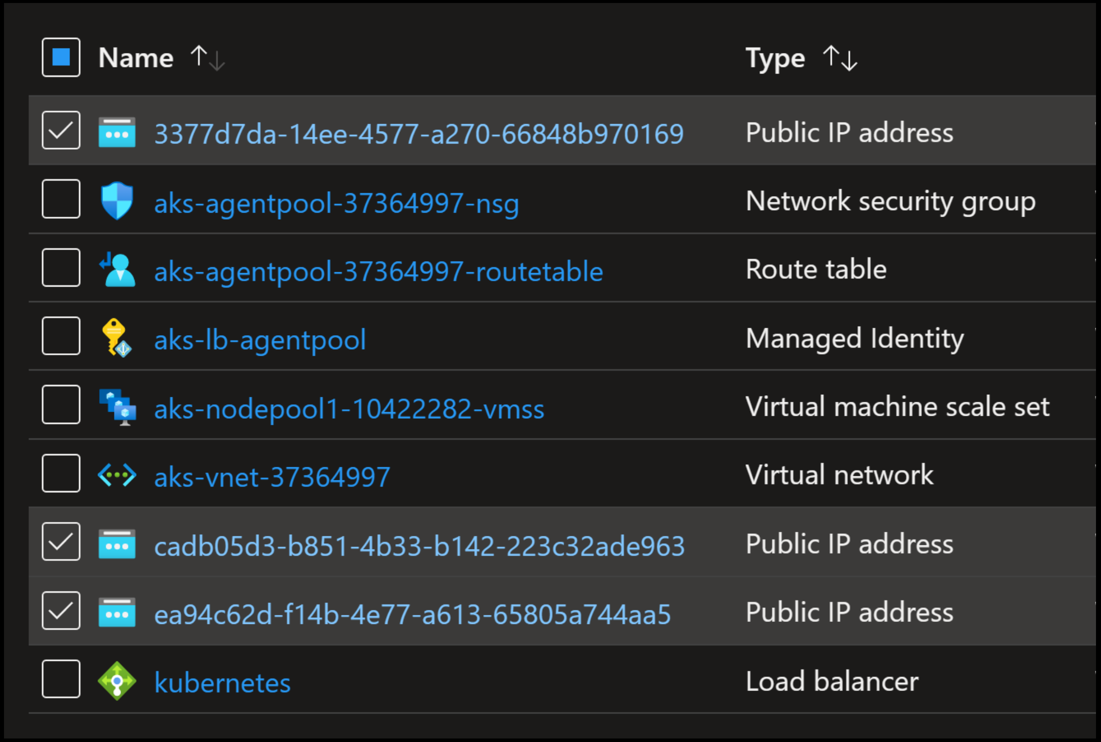
We can scale the number of managed outbound public IPs.

Each IP address provides 64k ephemeral ports to use as SNAT ports.

```
az aks update -g rg-aks-lb -n aks-lb `
   --load-balancer-managed-outbound-ip-count 3
```

But still the free pre-allocated IPs are not reused by other VMs.

Might be acceptable at a certain limit



| ■ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| ☑ | 3377d7da-14ee-4577-a270-66848b970169 | Public IP address |
| □ | aks-agentpool-37364997-nsg | Network security group |
| □ | aks-agentpool-37364997-routetable | Route table |
| □ | aks-lb-agentpool | Managed Identity |
| □ | aks-nodepool1-10422282-vmss | Virtual machine scale set |
| □ | aks-vnet-37364997 | Virtual network |
| ☑ | cadb05d3-b851-4b33-b142-223c32ade963 | Public IP address |
| ☑ | ea94c62d-f14b-4e77-a613-65805a744aa5 | Public IP address |
| □ | kubernetes | Load balancer |

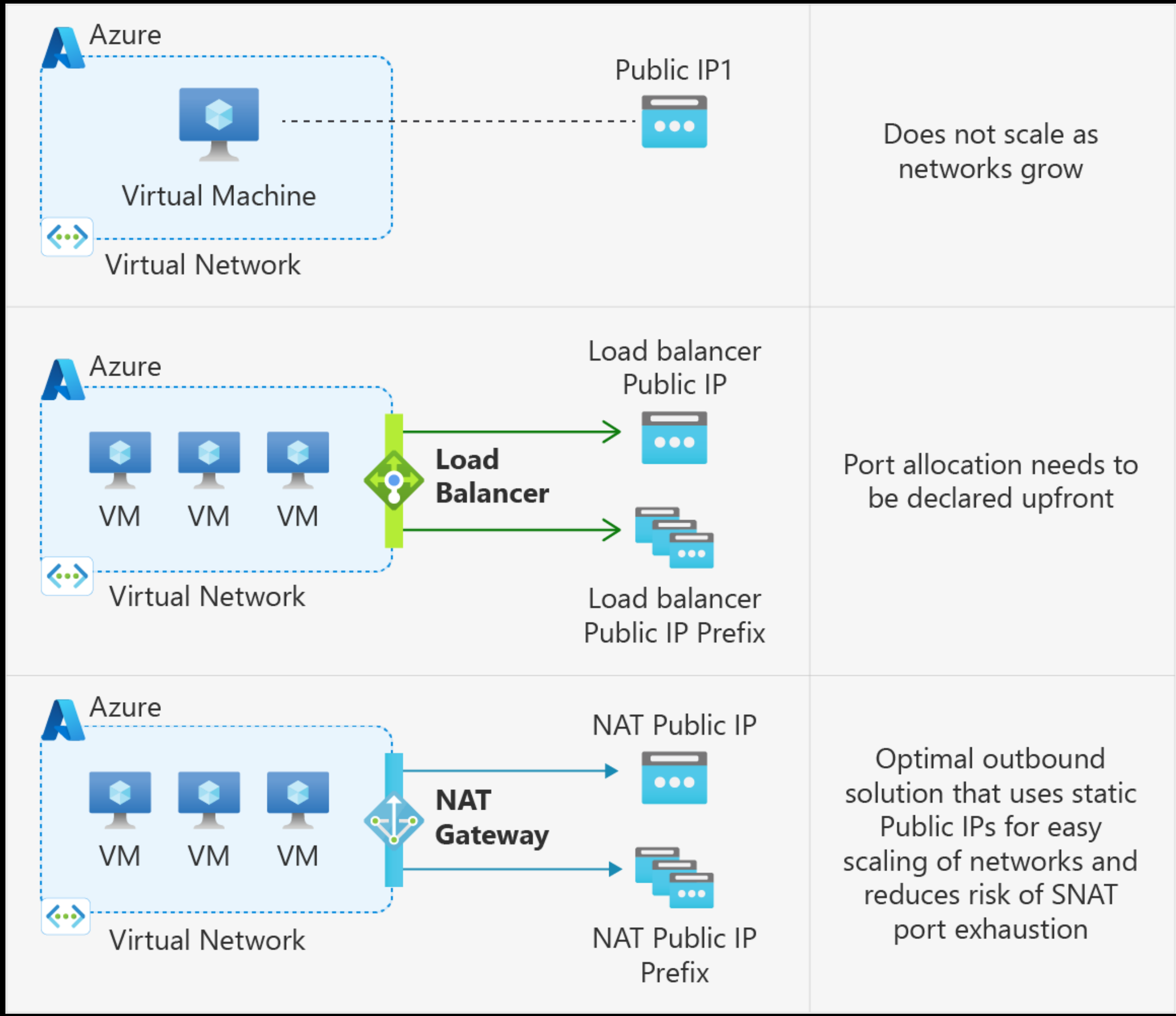
# What is Azure NAT Gateway ? (solution 2)

Virtual Network NAT is a fully managed and highly resilient Network Address Translation (NAT) service.

It simplifies outbound Internet connectivity for virtual networks.

It acts “like” a Load Balancer for outbound traffic.

And it reduces the risk of SNAT port exhaustion.



Does not scale as networks grow

Port allocation needs to be declared upfront

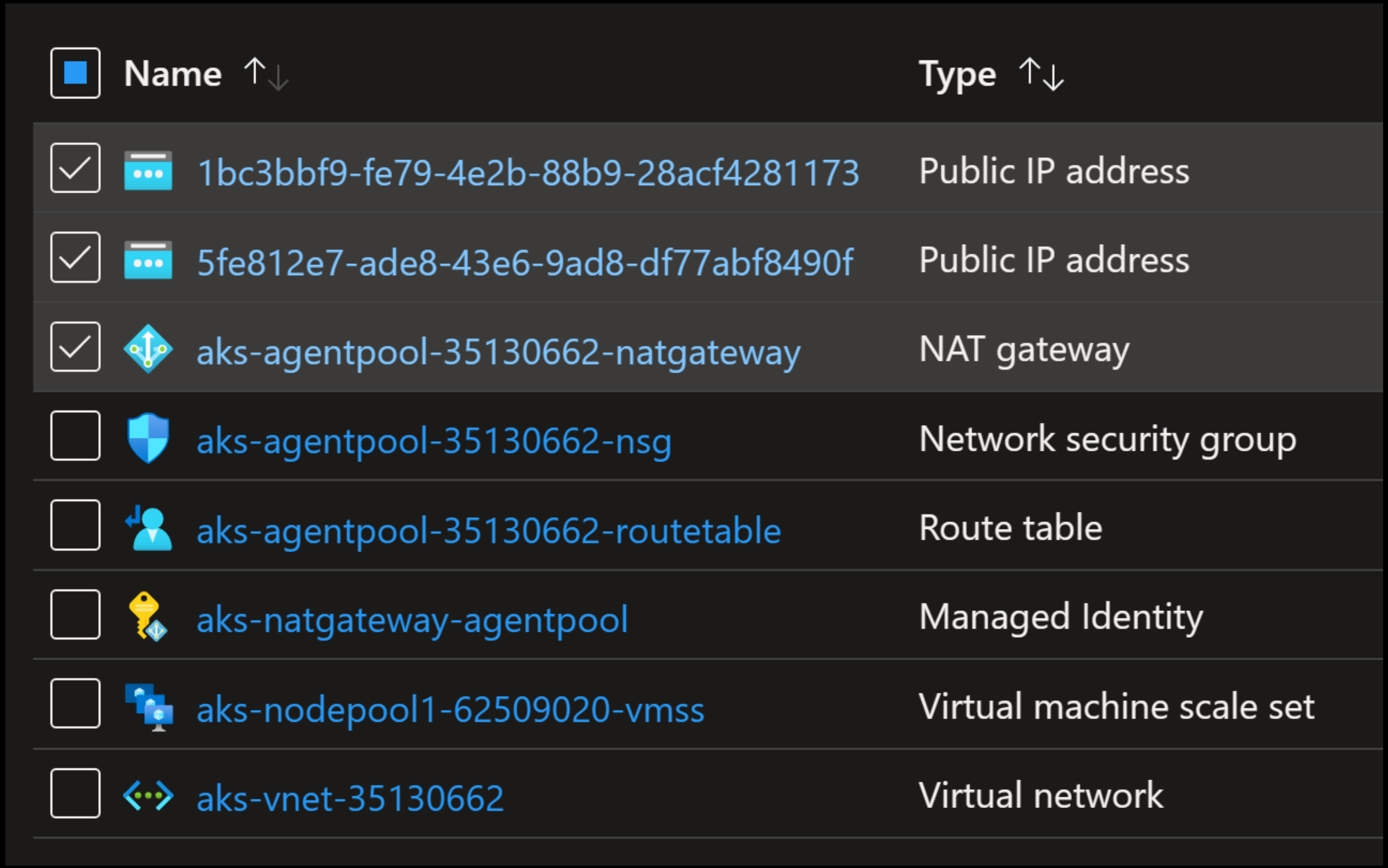
Optimal outbound solution that uses static Public IPs for easy scaling of networks and reduces risk of SNAT port exhaustion

# AKS OutboundType ManagedNATGateway

```
az aks create -g rg-aks-natgateway -n aks-natgateway `
   --outbound-type managedNATGateway `
   --nat-gateway-managed-outbound-ip-count 2 `
   --nat-gateway-idle-timeout 4
```

NAT Gateway and Public IPs are created.

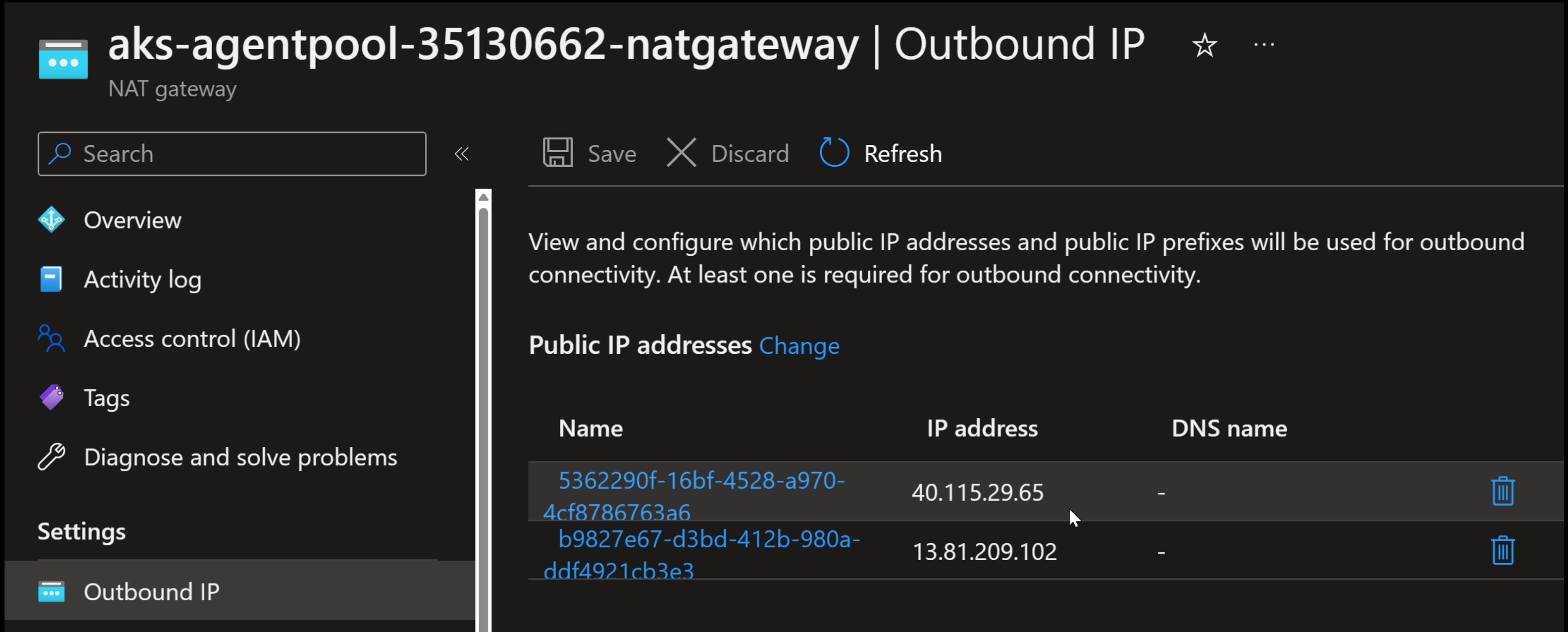
There are no Load Balancer.



| ■ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| ☑ | 1bc3bbf9-fe79-4e2b-88b9-28acf4281173 | Public IP address |
| ☑ | 5fe812e7-ade8-43e6-9ad8-df77abf8490f | Public IP address |
| ☑ | aks-agentpool-35130662-natgateway | NAT gateway |
| □ | aks-agentpool-35130662-nsg | Network security group |
| □ | aks-agentpool-35130662-routetable | Route table |
| □ | aks-natgateway-agentpool | Managed Identity |
| □ | aks-nodepool1-62509020-vmss | Virtual machine scale set |
| □ | aks-vnet-35130662 | Virtual network |

# Pods egress through NAT Gateway public IPs

```
kubectl run nginx --image=nginx
kubectl exec nginx -it -- /bin/bash
root@nginx:/# curl https://ifconfig.me
13.81.209.102
root@nginx:/# curl https://ifconfig.me
40.115.29.65
```



aks-agentpool-35130662-natgateway | Outbound IP

NAT gateway

View and configure which public IP addresses and public IP prefixes will be used for outbound connectivity. At least one is required for outbound connectivity.

**Public IP addresses** Change

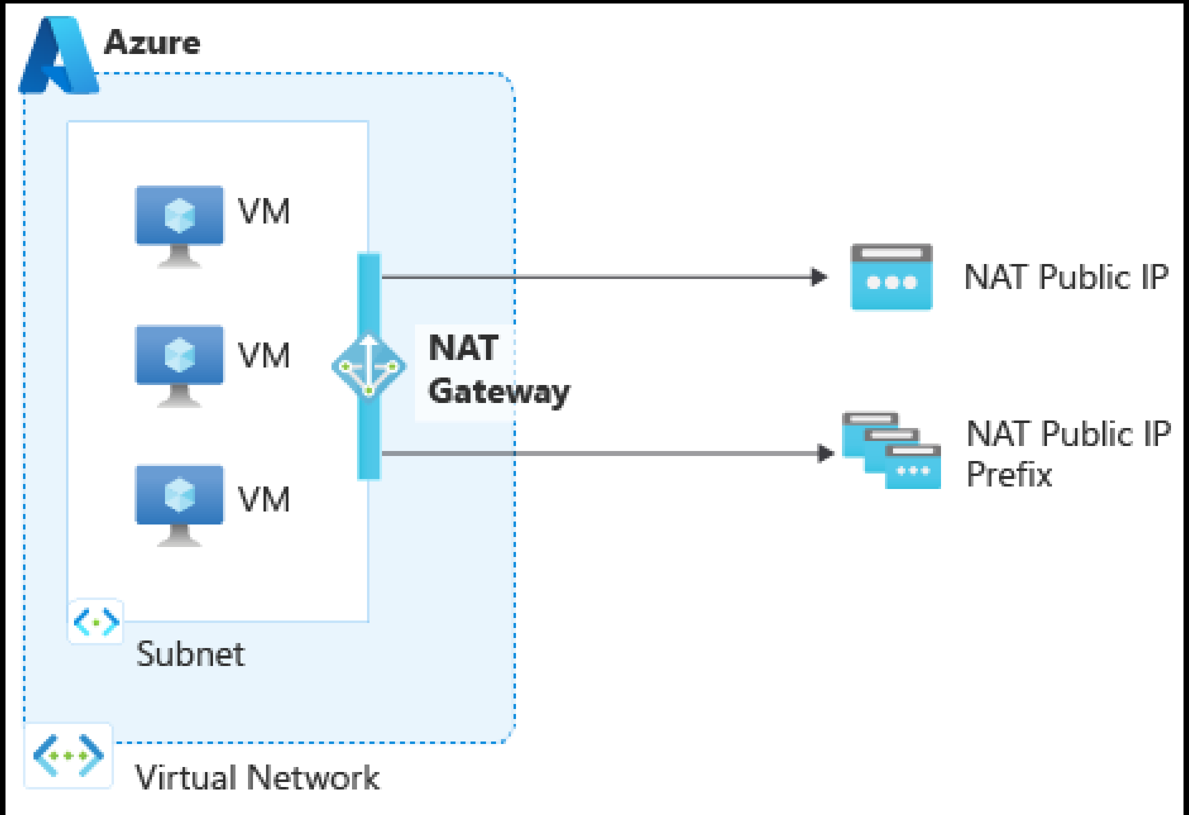
| Name | IP address | DNS name |
|---|---|---|
| 5362290f-16bf-4528-a970-4cf8786763a6 | 40.115.29.65 | - |
| b9827e67-d3bd-412b-980a-ddf4921cb3e3 | 13.81.209.102 | - |

# AKS OutboundType NATGateway

NAT Gateway could have 1 to 16 public IPs or a public IP Prefix.

Each IP address provides 64k SNAT ports ephemeral ports to use as SNAT ports.

64k SNAT ports * 16 IPs = 1,024,000 (~1 million) max SNAT ports.

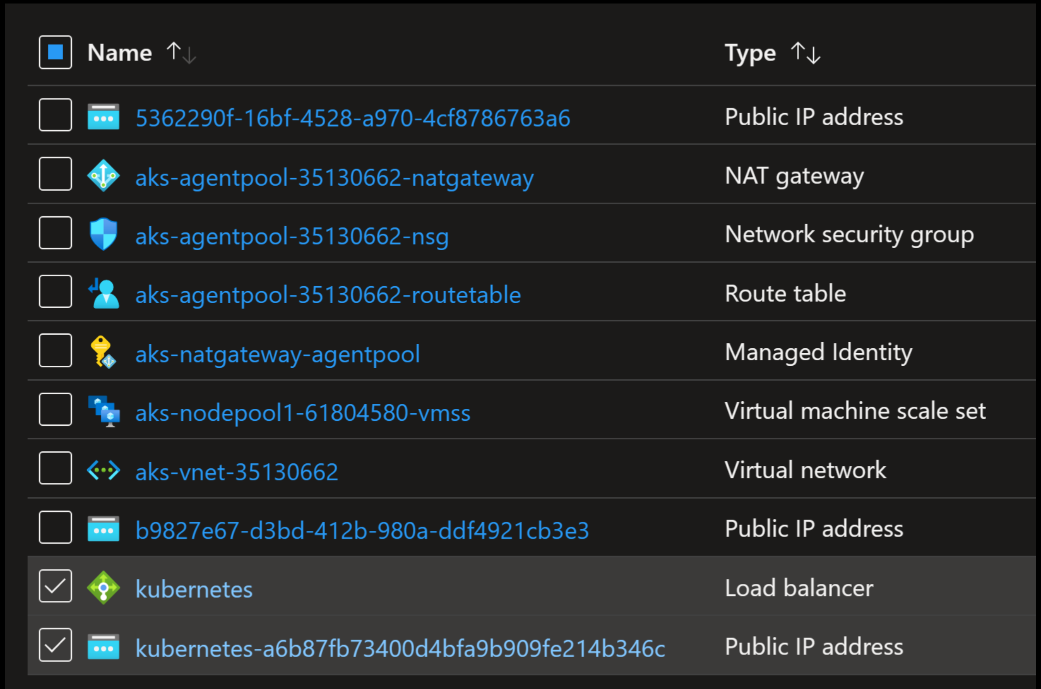


# Public service will create Load Balancer

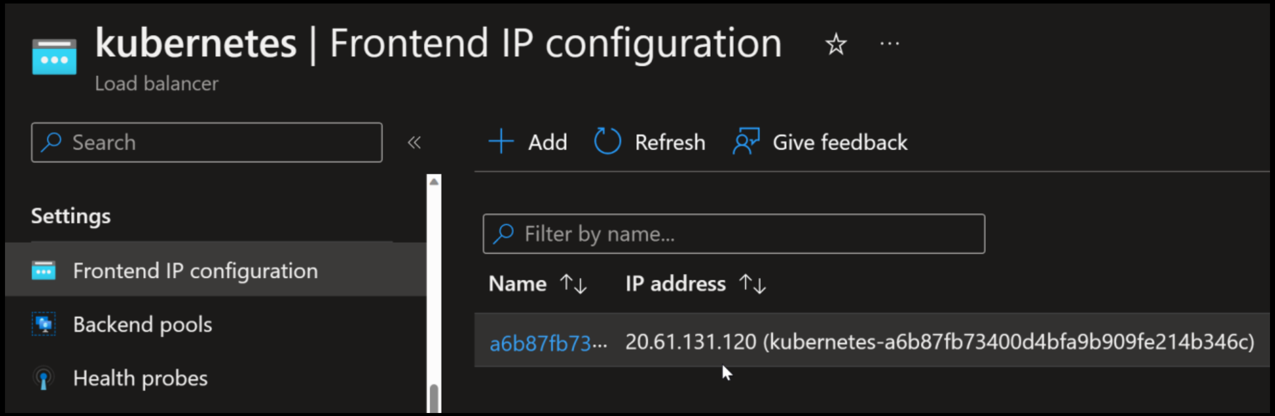
If we create a service of type LoadBalancer, AKS will create a new Load Balancer & public IP.

```
kubectl expose deployment nginx --name nginx --port=80 --type LoadBalancer

kubectl get svc
NAME        TYPE          CLUSTER-IP     EXTERNAL-IP    PORT(S)       AGE
kubernetes  ClusterIP     10.0.0.1       <none>         443/TCP       53m
nginx       LoadBalancer  10.0.138.225   20.61.131.120  80:31702/TCP  3m21s
```



| ■ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| ☐ | 5362290f-16bf-4528-a970-4cf8786763a6 | Public IP address |
| ☐ | aks-agentpool-35130662-natgateway | NAT gateway |
| ☐ | aks-agentpool-35130662-nsg | Network security group |
| ☐ | aks-agentpool-35130662-routetable | Route table |
| ☐ | aks-natgateway-agentpool | Managed Identity |
| ☐ | aks-nodepool1-61804580-vmss | Virtual machine scale set |
| ☐ | aks-vnet-35130662 | Virtual network |
| ☐ | b9827e67-d3bd-412b-980a-ddf4921cb3e3 | Public IP address |
| ☑ | kubernetes | Load balancer |
| ☑ | kubernetes-a6b87fb73400d4bfa9b909fe214b346c | Public IP address |



kubernetes | Frontend IP configuration

Load balancer

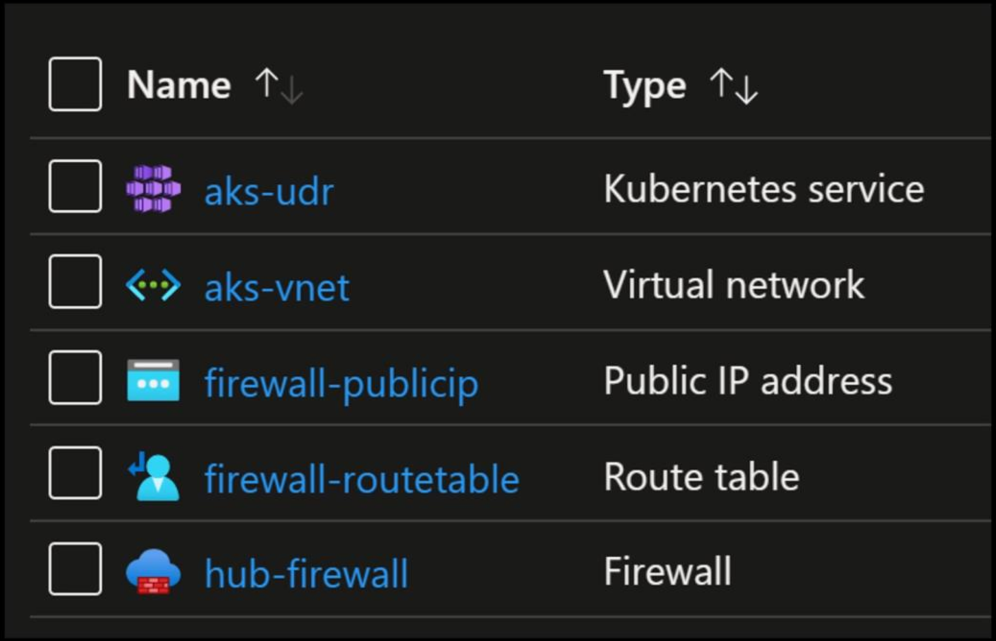
| Name ↑↓ | IP address ↑↓ |
|---|---|
| a6b87fb73... | 20.61.131.120 (kubernetes-a6b87fb73400d4bfa9b909fe214b346c) |

# AKS OutboundType UserDefinedRouting

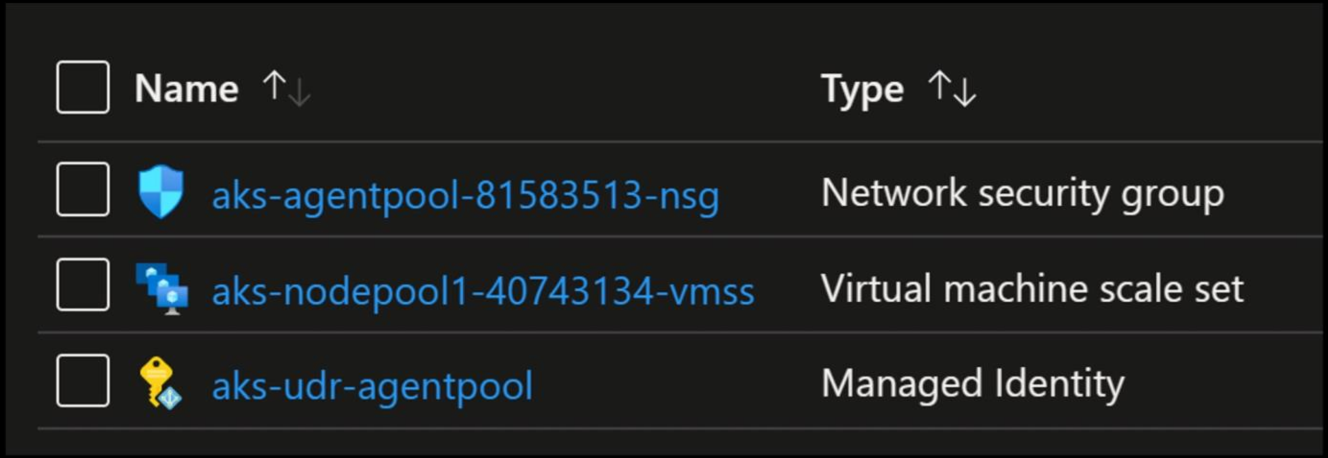
Useful when we want to filter and control AKS egress traffic, through a Firewall/NVA.
Widely used by enterprises adopting Hub & Spoke and Azure Landing Zones.

Creating route table, firewall, vnet, public IP, etc.

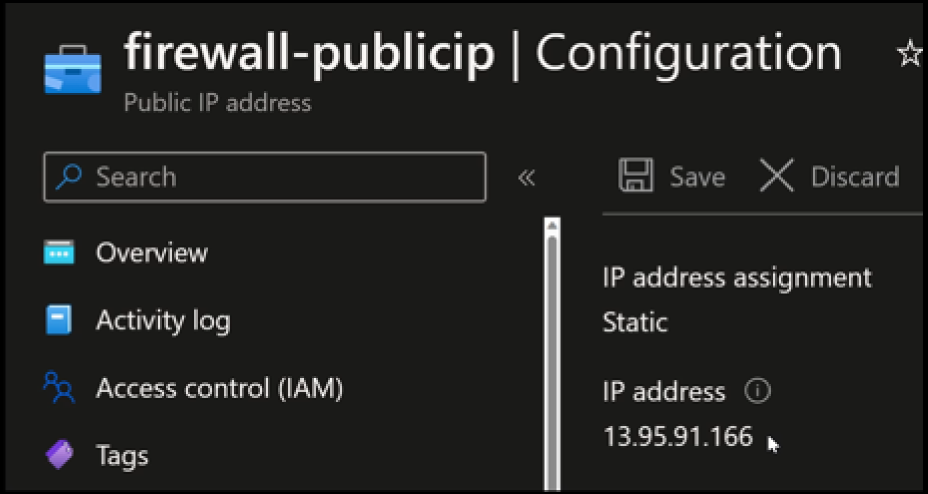
```
az aks create -g $RG -n $AKSNAME -l $LOC `
   --node-count 3 `
   --network-plugin azure `
   --outbound-type userDefinedRouting `
   --vnet-subnet-id $SUBNETID
```



| □ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| □ | aks-udr | Kubernetes service |
| □ | aks-vnet | Virtual network |
| □ | firewall-publicip | Public IP address |
| □ | firewall-routetable | Route table |
| □ | hub-firewall | Firewall |



| □ | Name ↑↓ | Type ↑↓ |
|---|---|---|
| □ | aks-agentpool-81583513-nsg | Network security group |
| □ | aks-nodepool1-40743134-vmss | Virtual machine scale set |
| □ | aks-udr-agentpool | Managed Identity |



firewall-publicip | Configuration ☆

Public IP address

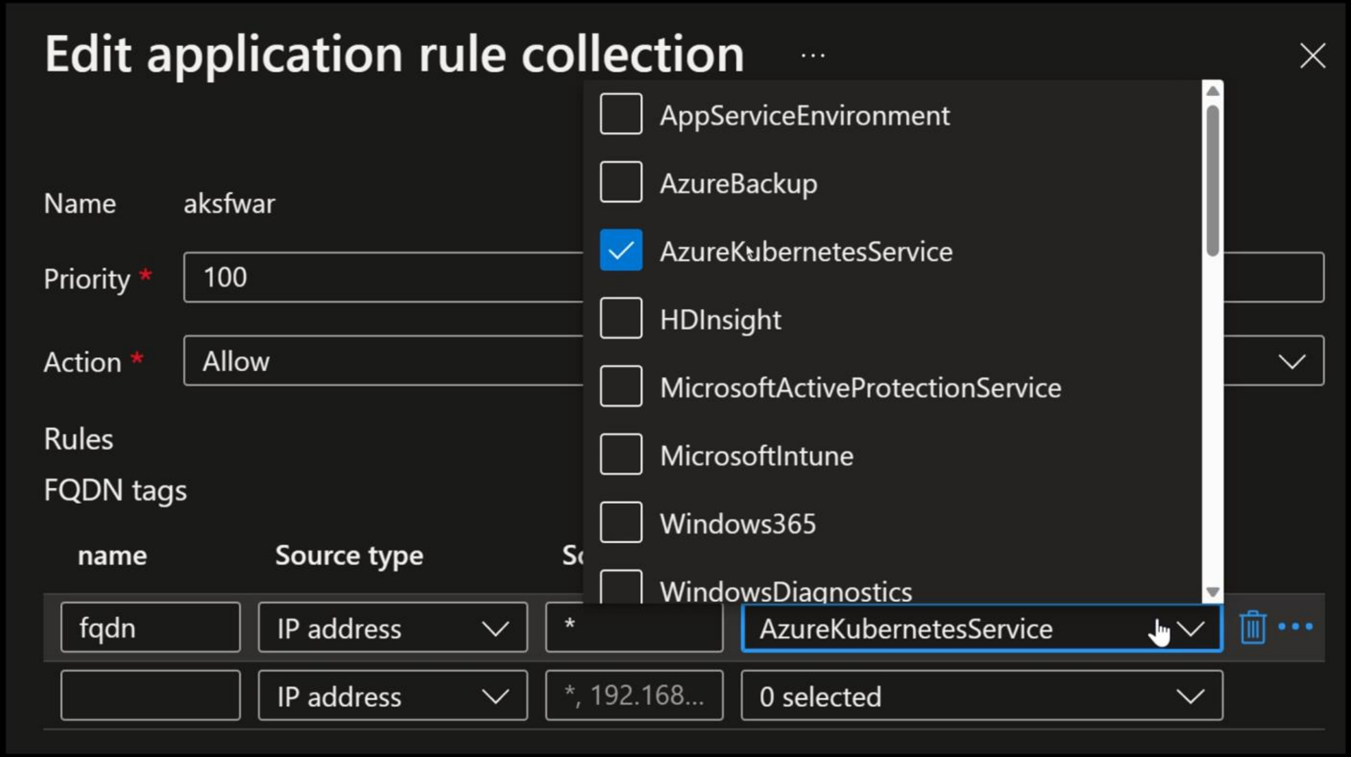
IP address assignment
Static

IP address ⓘ
13.95.91.166

# AKS OutboundType UserDefinedRouting

Adding FQDN Tag AzureKubernetesService to Firewall.
This allows AKS to access all the required services like OS updates, MCR, Azure Rest API, control plane…

```
az network firewall application-rule create -g $RG -f $FWNAME `
   --fqdn-tags "AzureKubernetesService" `
   --protocols 'http=80' 'https=443’ `
   --collection-name 'aksfwar’ `
   -n 'fqdn’ `
   --source-addresses '*’ `
   --action allow `
   --priority 100
```



## Edit application rule collection

Name aksfwar

Priority * 100

Action * Allow

Rules

FQDN tags

- □ AppServiceEnvironment
- □ AzureBackup
- ☑ AzureKubernetesService
- □ HDInsight
- □ MicrosoftActiveProtectionService
- □ MicrosoftIntune
- □ Windows365
- □ WindowsDiagnostics

| name | Source type | So |  |
|---|---|---|---|
| fqdn | IP address | * | AzureKubernetesService |
|  | IP address | *, 192.168... | 0 selected |

# Verify pod's egress traffic

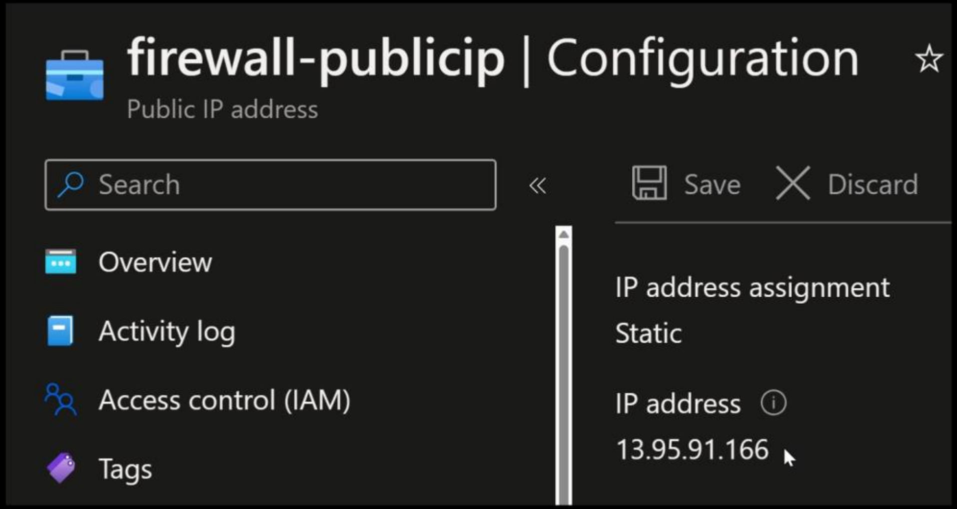
```
kubectl run nginx --image=nginx
pod/nginx created

kubectl get pods
NAME    READY   STATUS         RESTARTS   AGE
nginx   0/1     ErrImagePull   0          8s

az network firewall application-rule create <other_args>
   --target-fqdns hub.docker.com registry-1.docker.io production.cloudflare.docker.com
auth.docker.io cdn.auth0.com login.docker.com ifconfig.me

kubectl get pods -w
NAME    READY   STATUS    RESTARTS   AGE
nginx   1/1     Running   0          21m

kubectl exec nginx -it -- /bin/bash
root@nginx:/# curl http://ifconfig.me
13.95.91.166
```



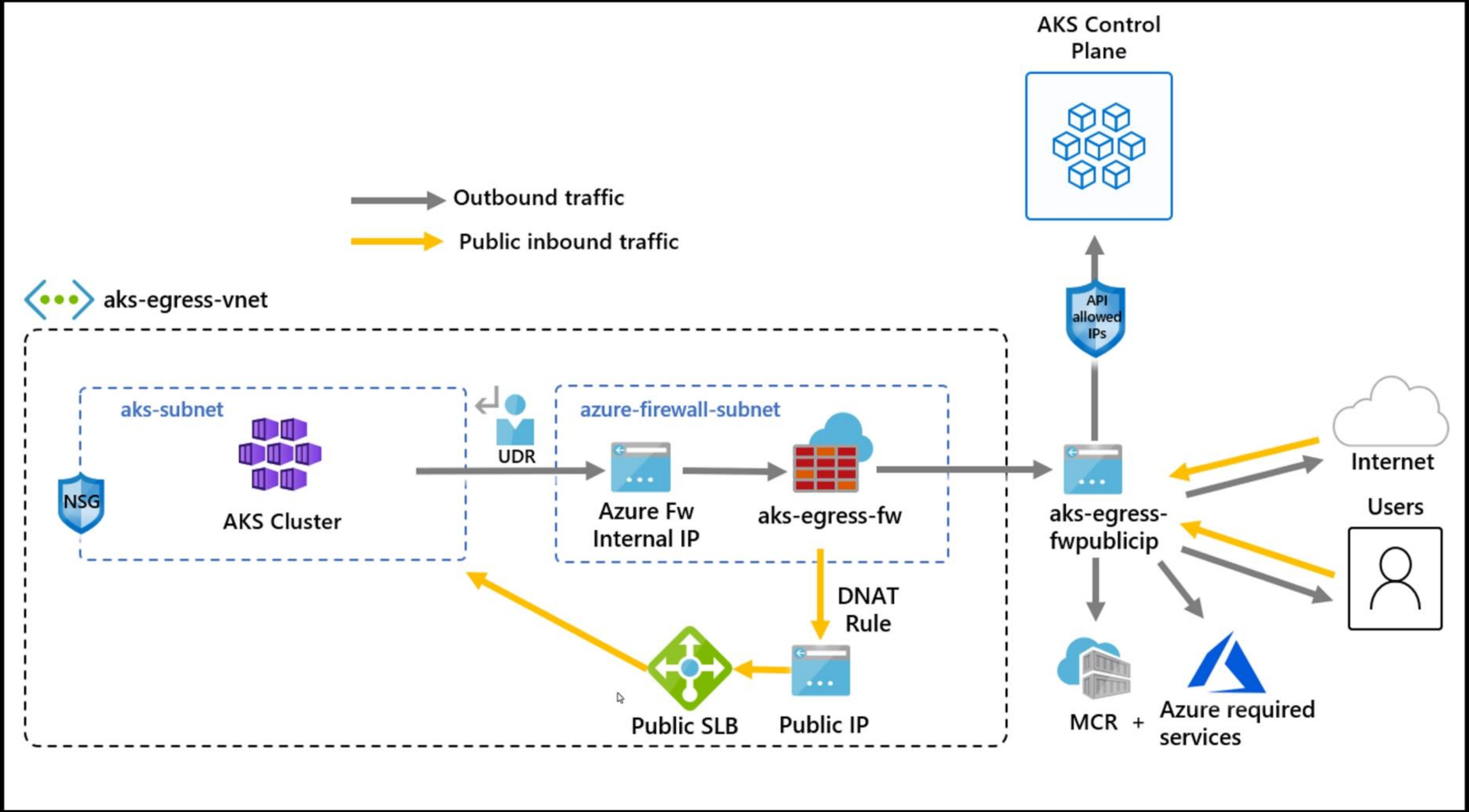
firewall-publicip | Configuration

Public IP address

IP address assignment
Static

IP address
13.95.91.166

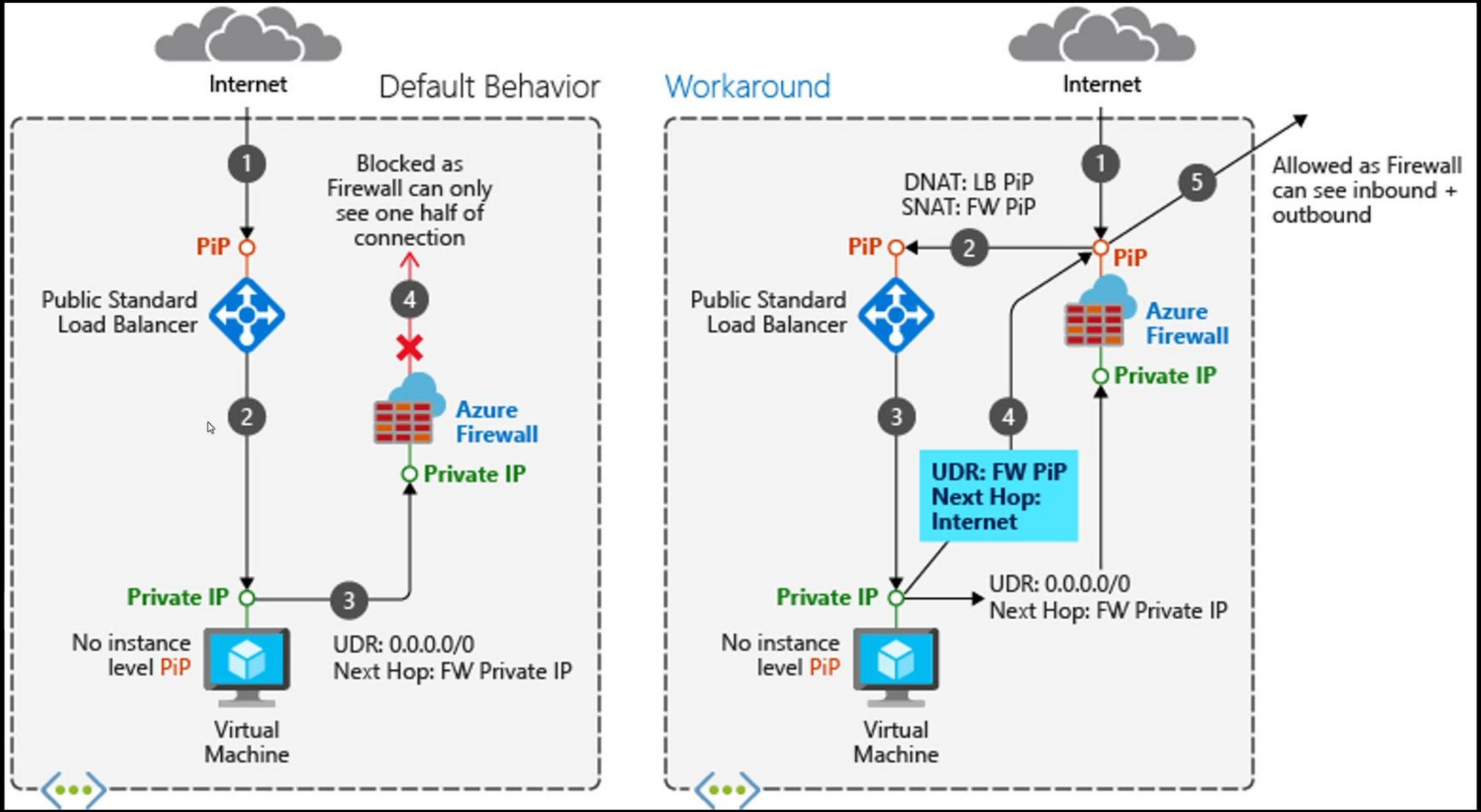
# Egress (and ingress) traffic through Firewall



https://learn.microsoft.com/en-us/azure/aks/limit-egress-traffic

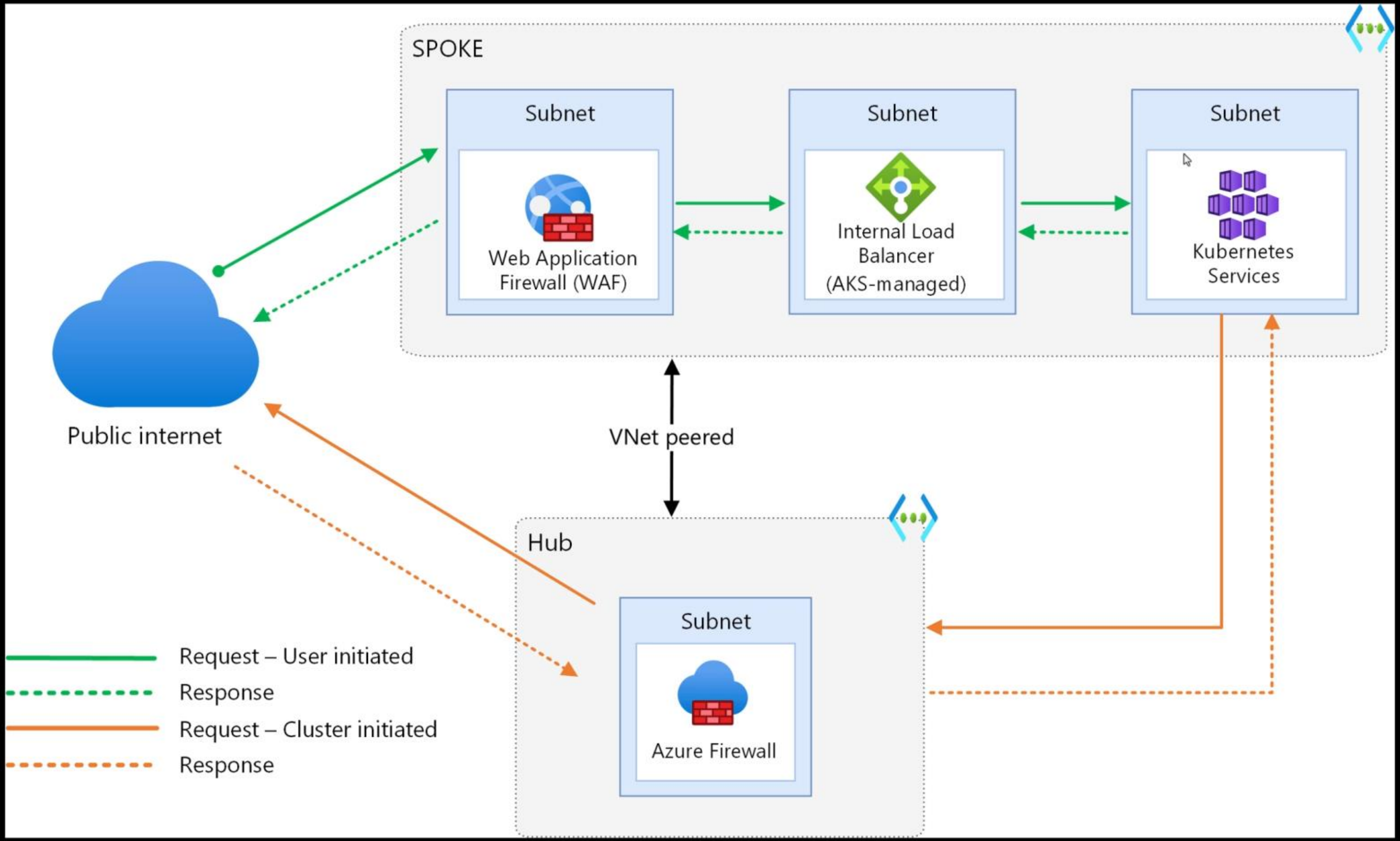
# Asymmetric routing issue for ingress w/ LB & Firewall

**Asymmetric routing issue:** https://learn.microsoft.com/en-us/azure/firewall/integrate-lb



# Ingress with App Gateway and egress with Firewall

**Application Gateway Ingress Controller (AGIC) won't have the asymmetric routing issue.**
**Because it is inside the AKS VNET, it injects its own private IP so the traffic will not be routed to the Firewall.**



# More resources

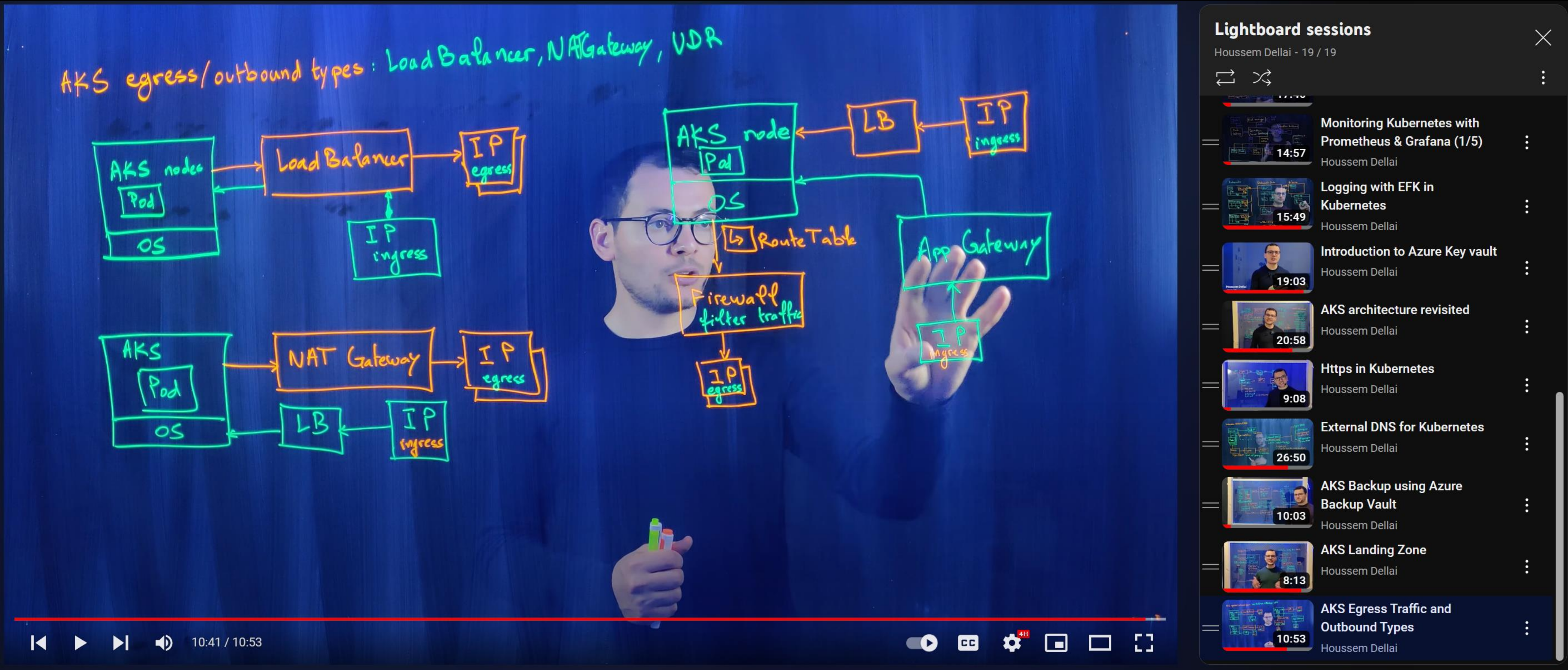
Filtering AKS egress traffic with Virtual WAN
https://blog.cloudtrooper.net/2023/01/10/filtering-aks-egress-traffic-with-virtual-wan/

Control egress traffic using Azure Firewall in AKS
https://learn.microsoft.com/en-us/azure/aks/limit-egress-traffic

Outbound network and FQDN rules for AKS
https://learn.microsoft.com/en-us/azure/aks/outbound-rules-control-egress

# More resources

youtube.com/watch?v=V2E1WNR-4KM&list=PLpbcUe4chE79jMdliWZi0QerwyJ7Zz18D&index=19



Lightboard sessions

Houssem Dellai - 19 / 19

- Monitoring Kubernetes with Prometheus & Grafana (1/5) — Houssem Dellai — 14:57
- Logging with EFK in Kubernetes — Houssem Dellai — 15:49
- Introduction to Azure Key vault — Houssem Dellai — 19:03
- AKS architecture revisited — Houssem Dellai — 20:58
- Https in Kubernetes — Houssem Dellai — 9:08
- External DNS for Kubernetes — Houssem Dellai — 26:50
- AKS Backup using Azure Backup Vault — Houssem Dellai — 10:03
- AKS Landing Zone — Houssem Dellai — 8:13
- AKS Egress Traffic and Outbound Types — Houssem Dellai — 10:53