Betrieb von Azure Kubernetes Service Erfahrungen aus der Praxis

Daniel Neumann
Staff Software Engineer – LeanIX
Microsoft MVP – Microsoft Azure
https://www.danielstechblog.io
@neumanndaniel

Experts Live Germany

Sessionüberlick

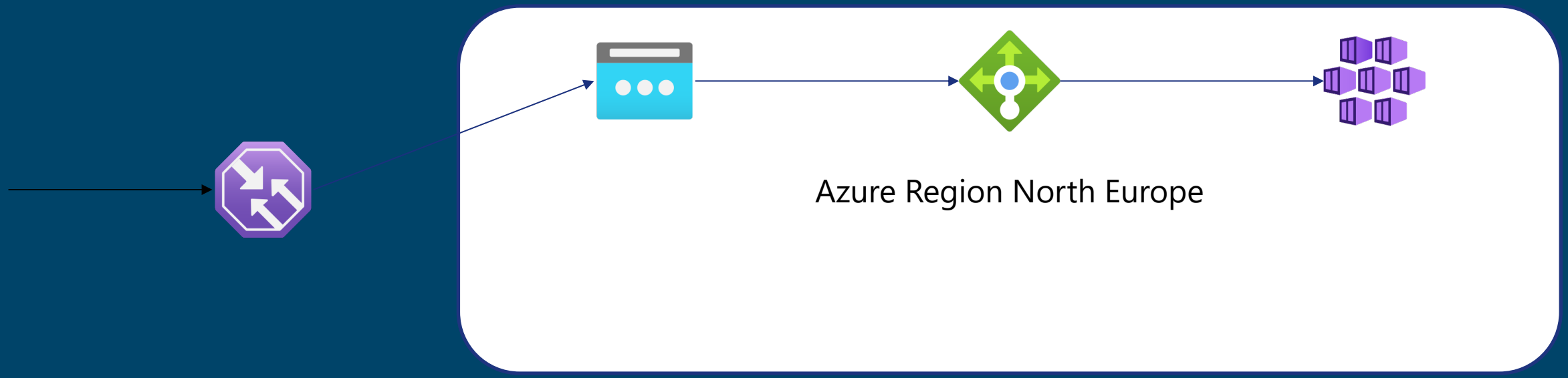
- Business Continuity und Desaster Recovery
- Hochverfügbarkeit
- Patch und Upgrade Management
- Governance, Monitoring und Alerting

Business Continuity und Desaster Recovery

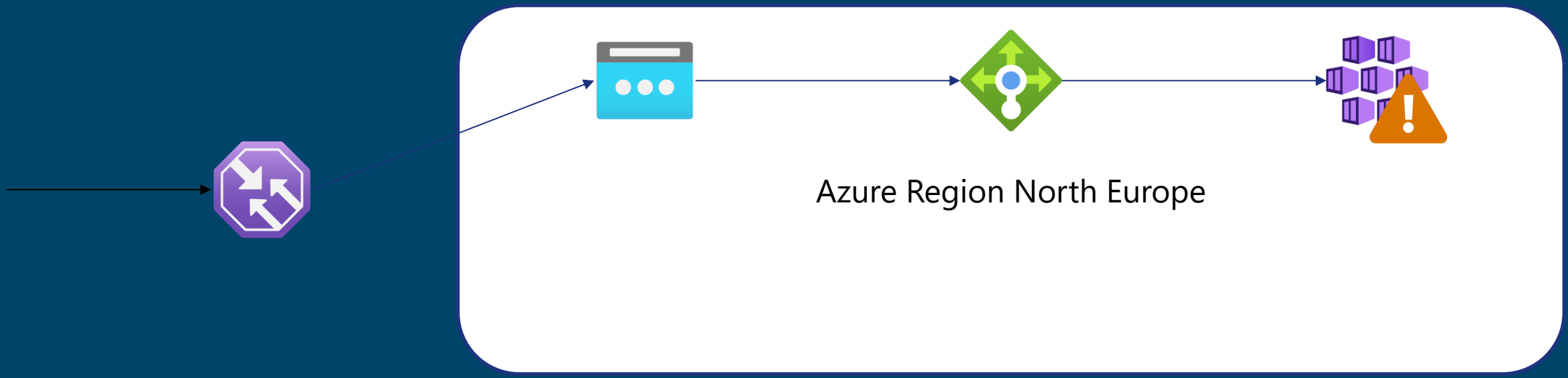
Business Continuity und Desaster Recovery

- DR Strategie
- Azure Traffic Manager als DNS Load Balancer
- Stateless Services –> PaaS Services
  - Azure Storage
  - Azure Database for PostgreSQL
  - Azure Cache for Redis
  - etc.

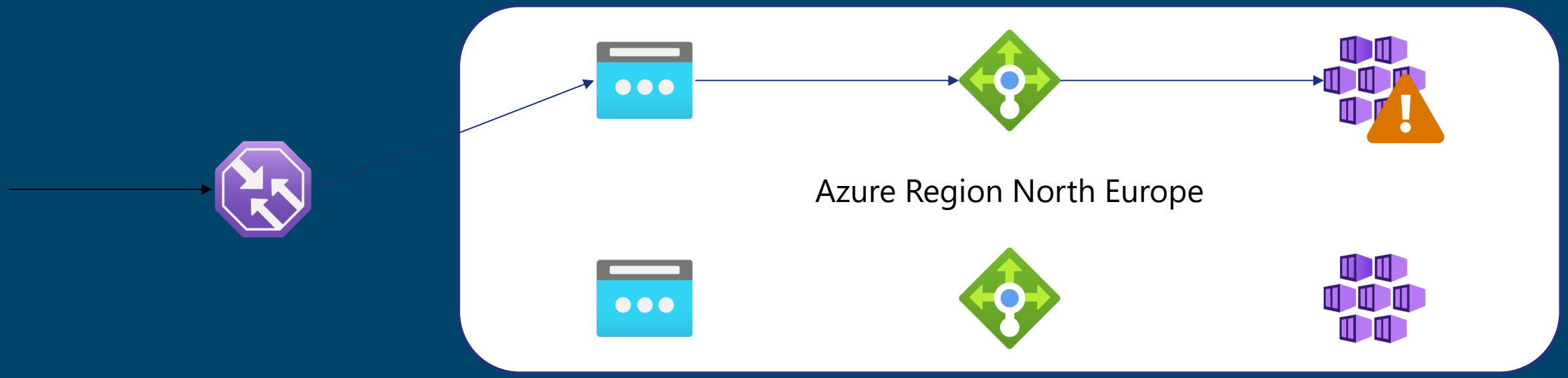
BCDR – Traffic Manager Architecture



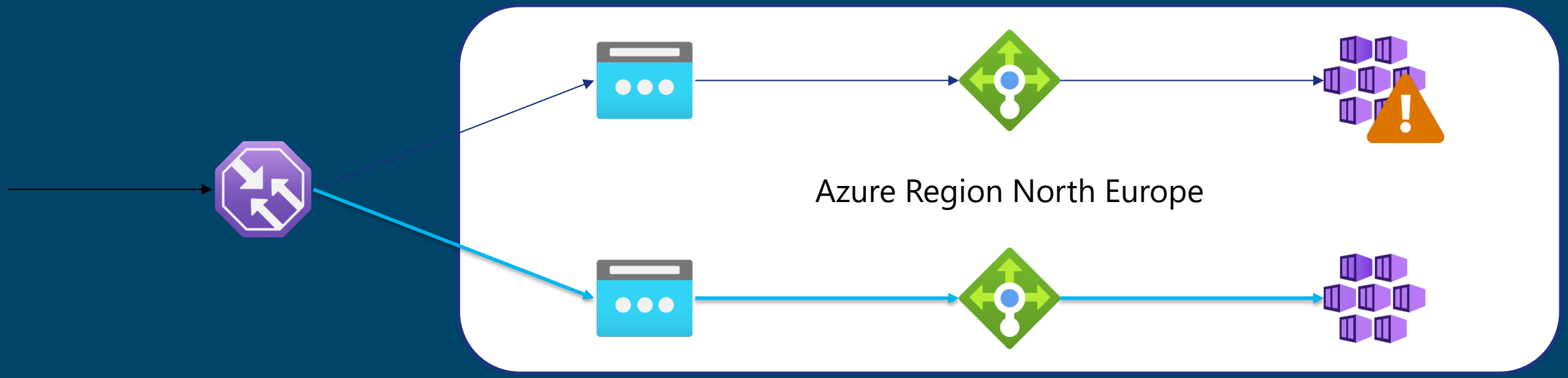
BCDR – Traffic Manager Architecture



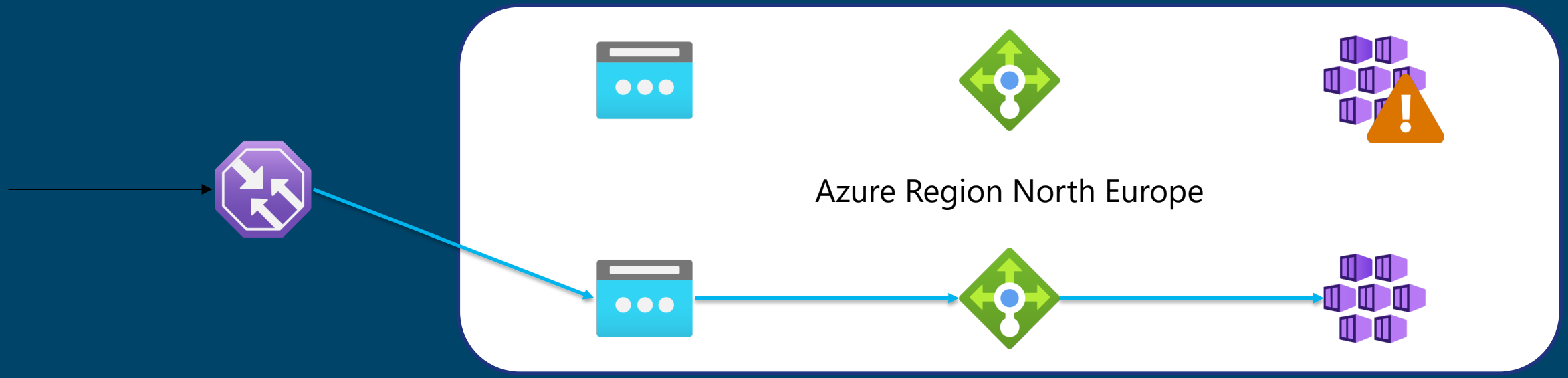
BCDR – Traffic Manager Architecture



BCDR – Traffic Manager Architecture



BCDR – Traffic Manager Architecture



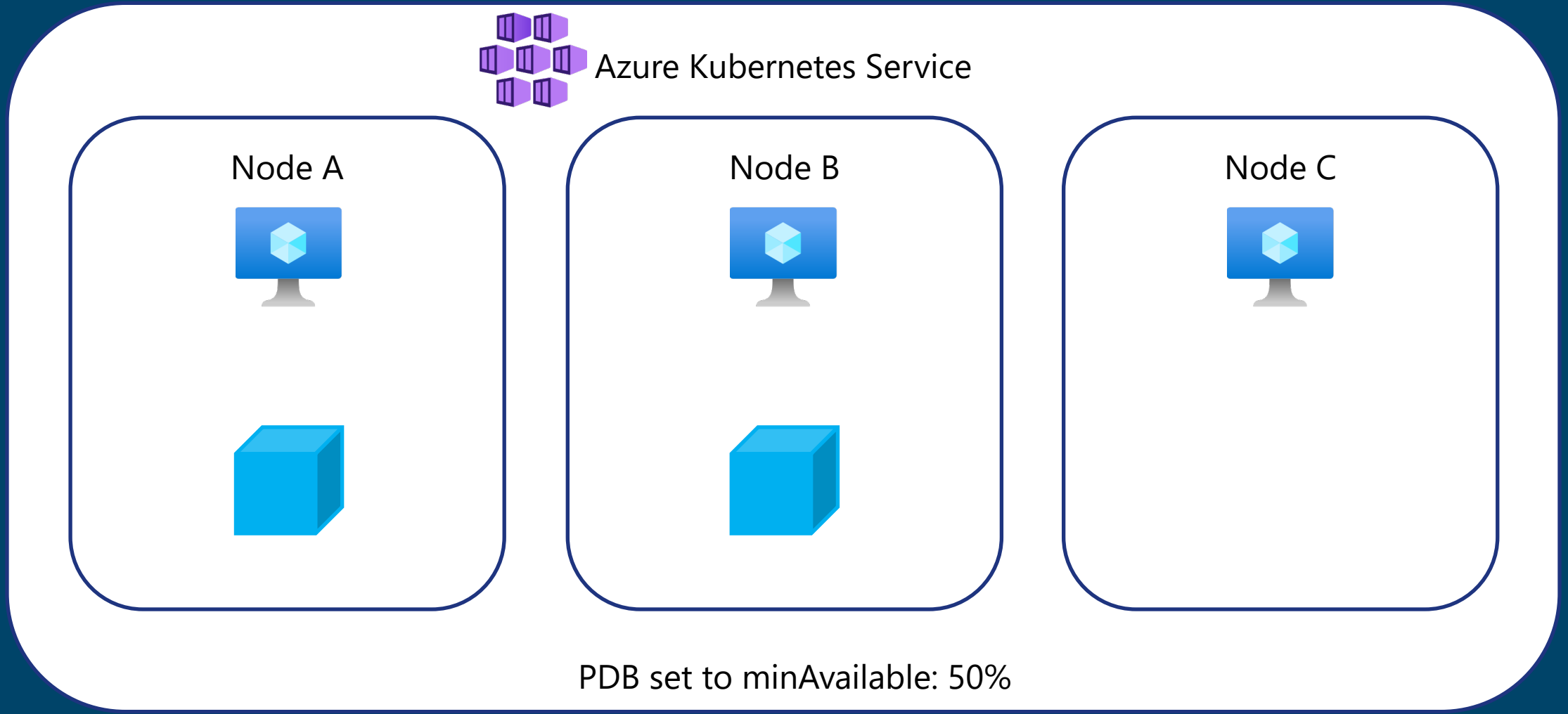
Demo – Traffic Manager

Hochverfügbarkeit

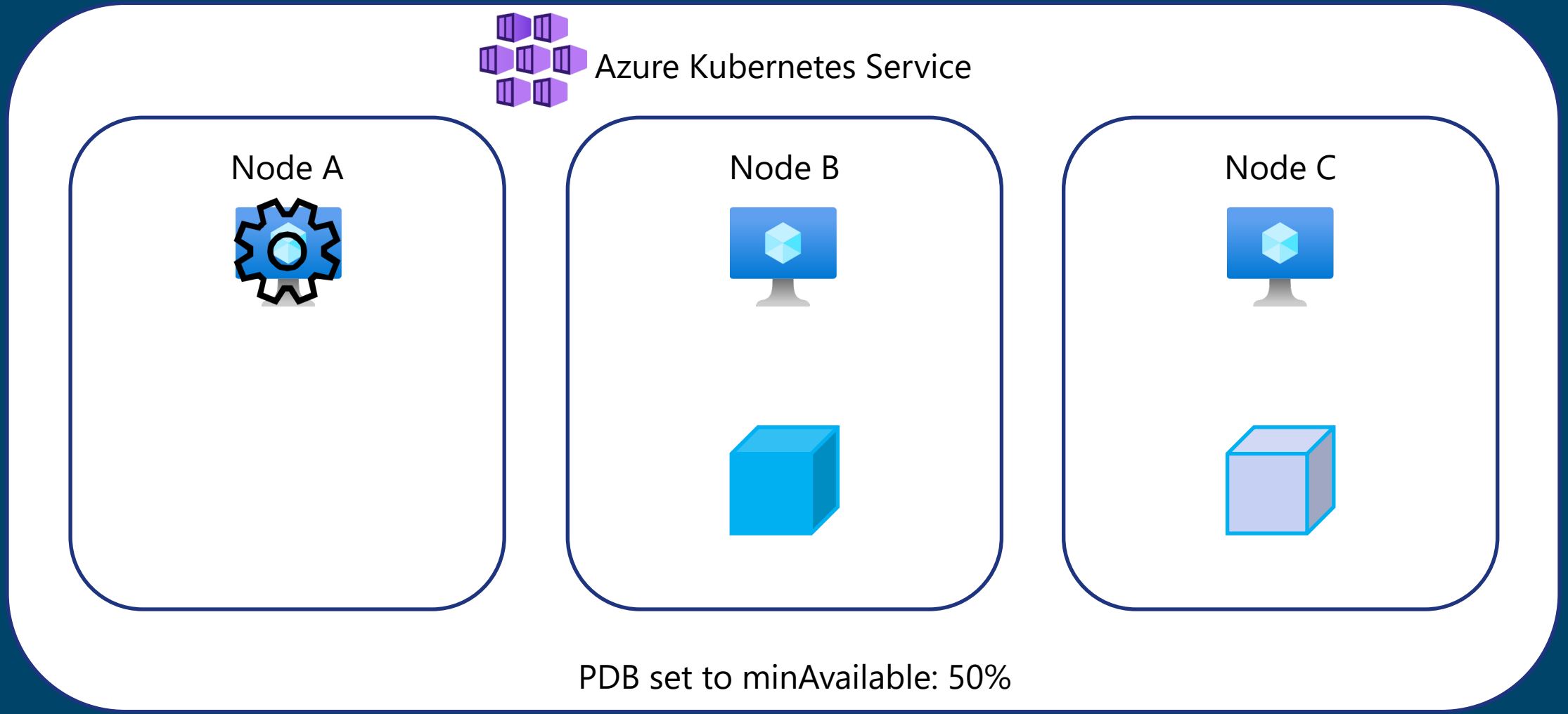
Hochverfügbarkeit

- Pod Disruption Budgets für Applikationen
  - Maintenance
- Pod Anti Affinity Einstellungen zur Verteilung der Applikation über Nodes / Availability Zones
  - Ausfälle
- Pod Topology Spread Constraints zur Verteilung der Applikation über Availability Zones
  - Kubernetes 1.19 oder höher
  - Ausfälle

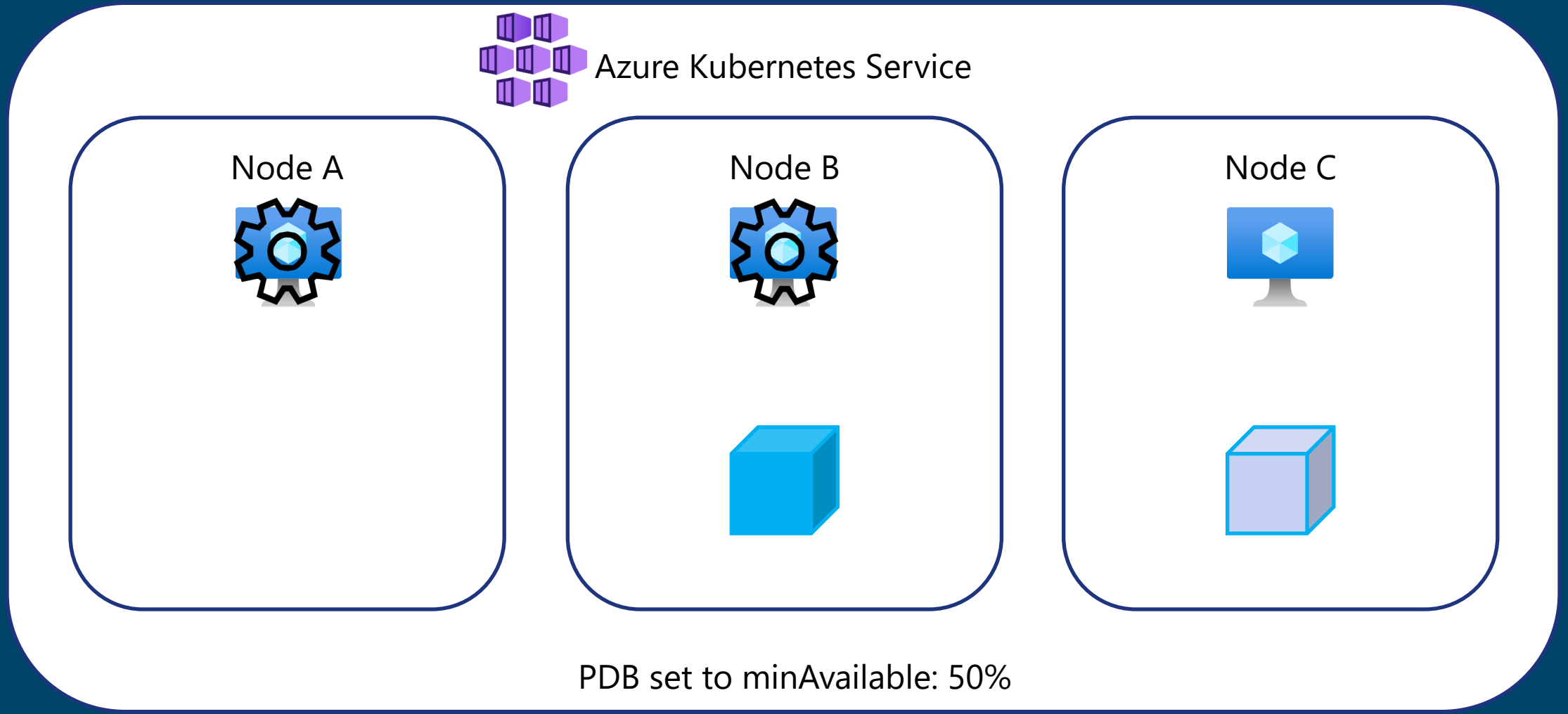
HA – Pod Disruption Budgets



HA – Pod Disruption Budgets



HA – Pod Disruption Budgets



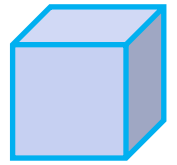
HA – Pod Disruption Budgets

Azure Kubernetes Service

Node A

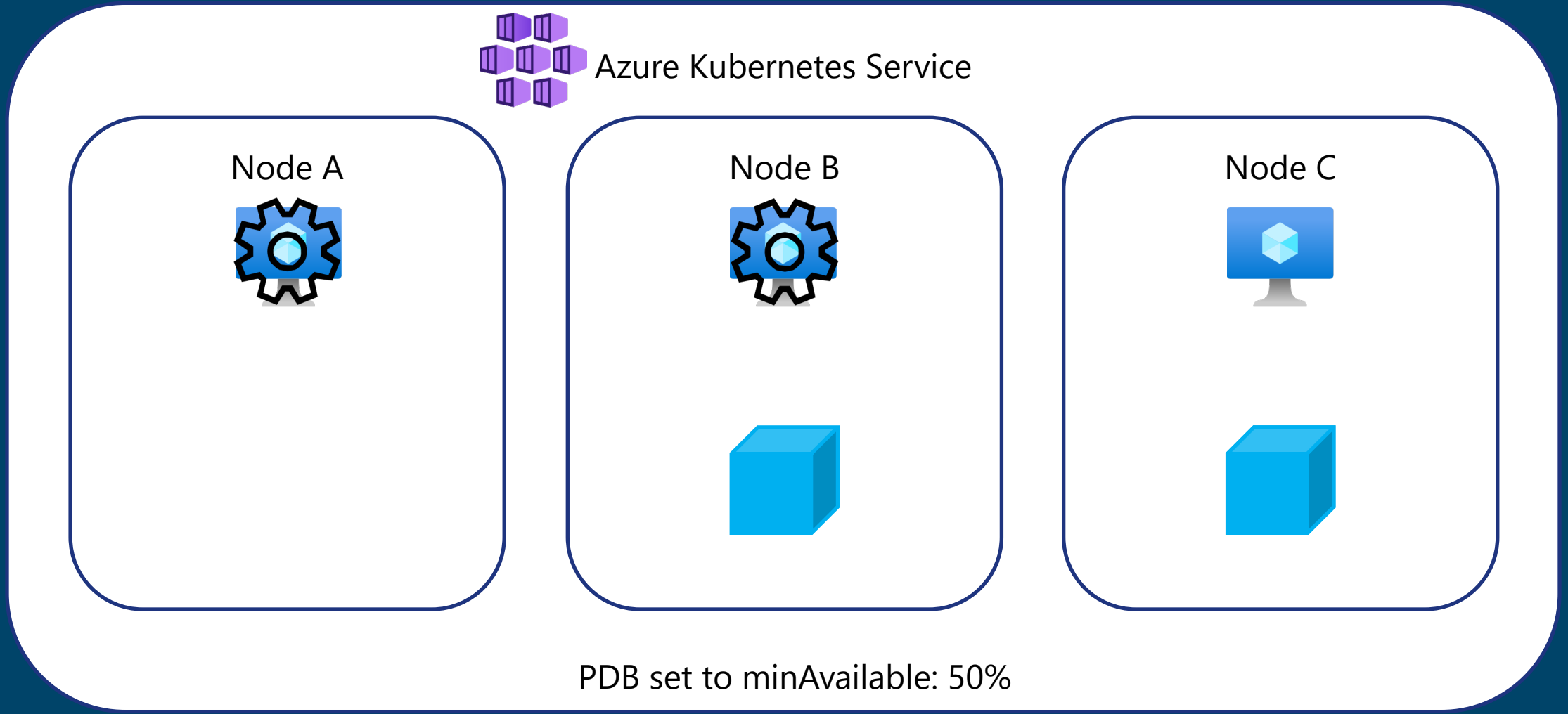
Node B

Node C



PDB set to minAvailable: 50%

HA – Pod Disruption Budgets

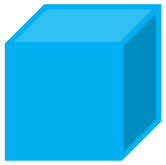


HA – Pod Topology Spread Constraints

Azure Kubernetes Service

Node A – Zone A

A-B: 2-1=1
A-C: 2-0=2



Node B – Zone B

B-A: 2-1=1
B-C: 2-0=2

Node C – Zone C

C-A: 0-1=-1
C-B: 0-1=-1

PTSC maxSkew: 1

Hochverfügbarkeit

- Kubernetes readiness probes
  - Stellen sicher, dass der Service Traffic empfängt, wenn er bereit ist

Cluster-Autoscaling

- Cluster Autoscaler
  - Einfaches Ersetzen von Nodes, die nicht von der Node Auto-Repair Funktion erkannt und repariert wurden
- Quota regelmäßig überprüfen
- AKS Uptime SLA
  - 99,95% SLA mit AZ (Availability Zones)
  - 99,9% SLA
  - 99,5% **SLO** im Free Tier

Service-Autoscaling

- KEDA für erweiterte Pod Scaling Szenarien
  - Standard HPA (Horizontal Pod Autoscaler) in Kubernetes unterstützt nur CPU und Memory

Patch und Upgrade Management

Azure Kubernetes Service up-to-date

- kured – Kubernetes Reboot Daemon
- Node OS Image Upgrade
  - Neue Images erscheinen alle 1-3 Wochen
- Automatisches Kubernetes Version Upgrade

Governance, Monitoring und Alerting

Governance mit Azure Policy

- Azure Policy für Kubernetes (Gatekeeper)
  - Governance -> Labels
  - Security -> Non-root Container
  - Compliance -> Genehmigte Images
  - CVEs -> CVE-2020-8554

Monitoring und Alerting

- AKS – Diagnostic Checks
- Azure Monitor – Container insights
  - Empfohlene Alerts
  - Log query Alerts
- AKS Control Plane Logs
  - Storage Account empfohlen
- Kubernetes Ressourcen im Azure Portal

Demo

SNAT Port Exhaustion

Governance, Monitoring und Alerting

Experts Live Germany

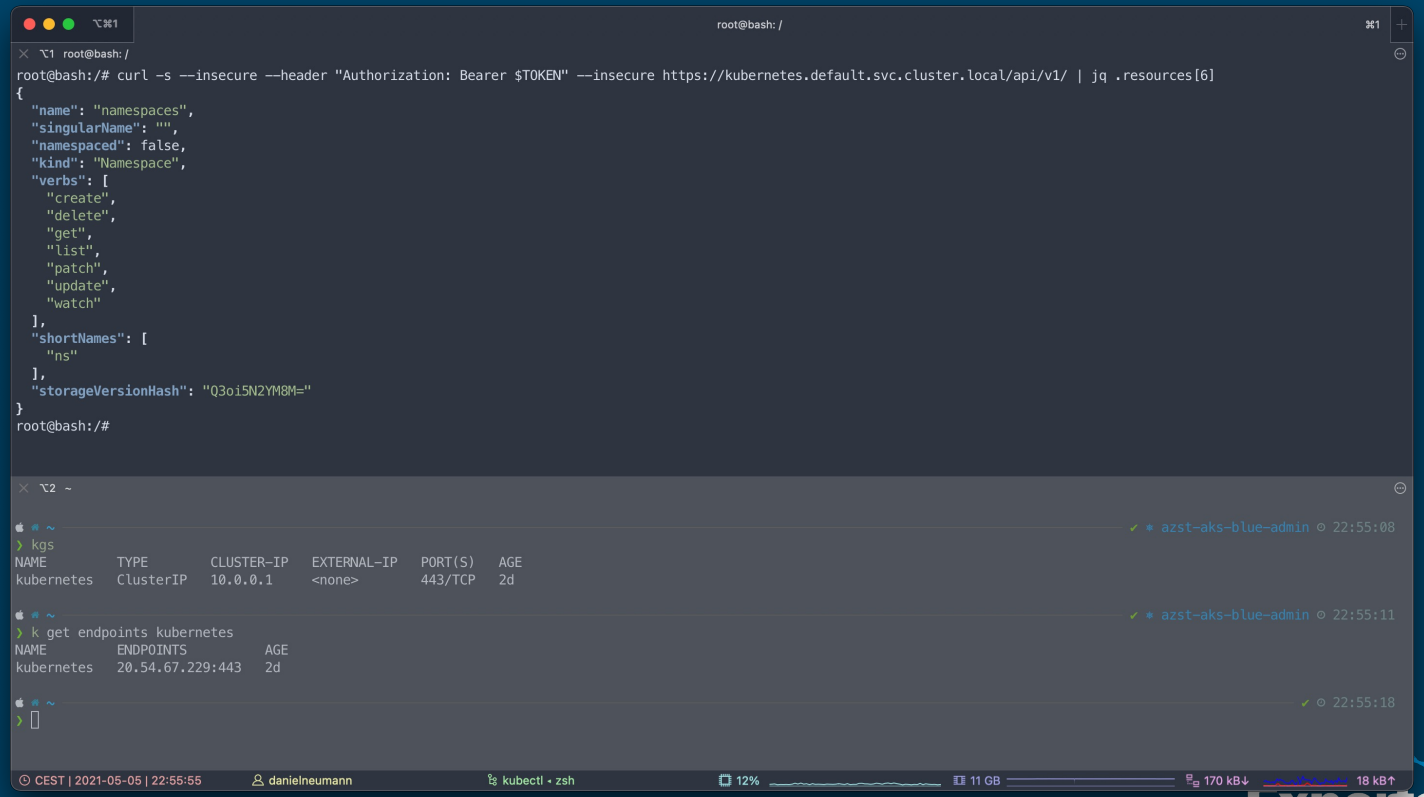
SNAT Verbindungen – Was zählt?

- Zugriff auf Azure PaaS ohne Private Link
  - Azure Storage
  - Azure Database for PostgreSQL
  - Azure Cache for Redis
  - etc.

- Zugriff auf externe Services

SNAT Verbindungen – Was zählt?

- AKS API Server Zugriff innerhalb des Clusters
  - kubernetes.default.svc.cluster.local



Behebung

1. Es müssen genügend Public IPs dem Load Balancer zugewiesen, es muss der Wert für die allokierten SNAT Ports per Node angepasst und der TCP Idle Rest auf 4 Minuten gesetzt werden
   - Der automatische Default hängt von der Clustergröße ab und fängt bei 1024 SNAT Ports an und endet bei 32 SNAT Ports pro Node
   - Es besteht ein Restrisiko für eine SNAT Port Exhaustion

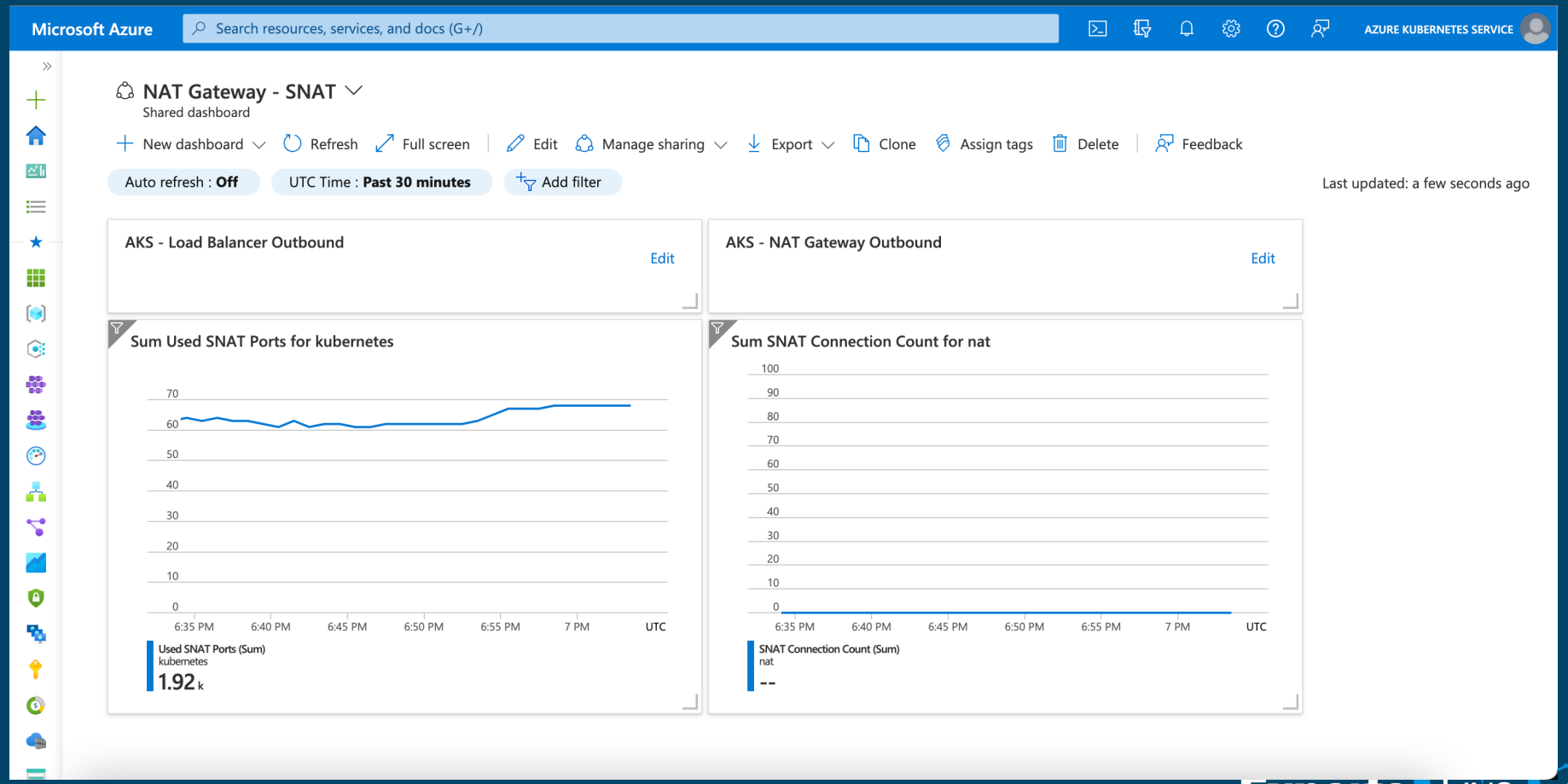
2. Azure Virtual Network NAT nutzen
   - Nicht den *outboundType* managedNATGateway oder userAssignedNATGateway in der AKS Konfiguration nutzen

Warum sollte ich bei *outboundType* loadBalancer bleiben?

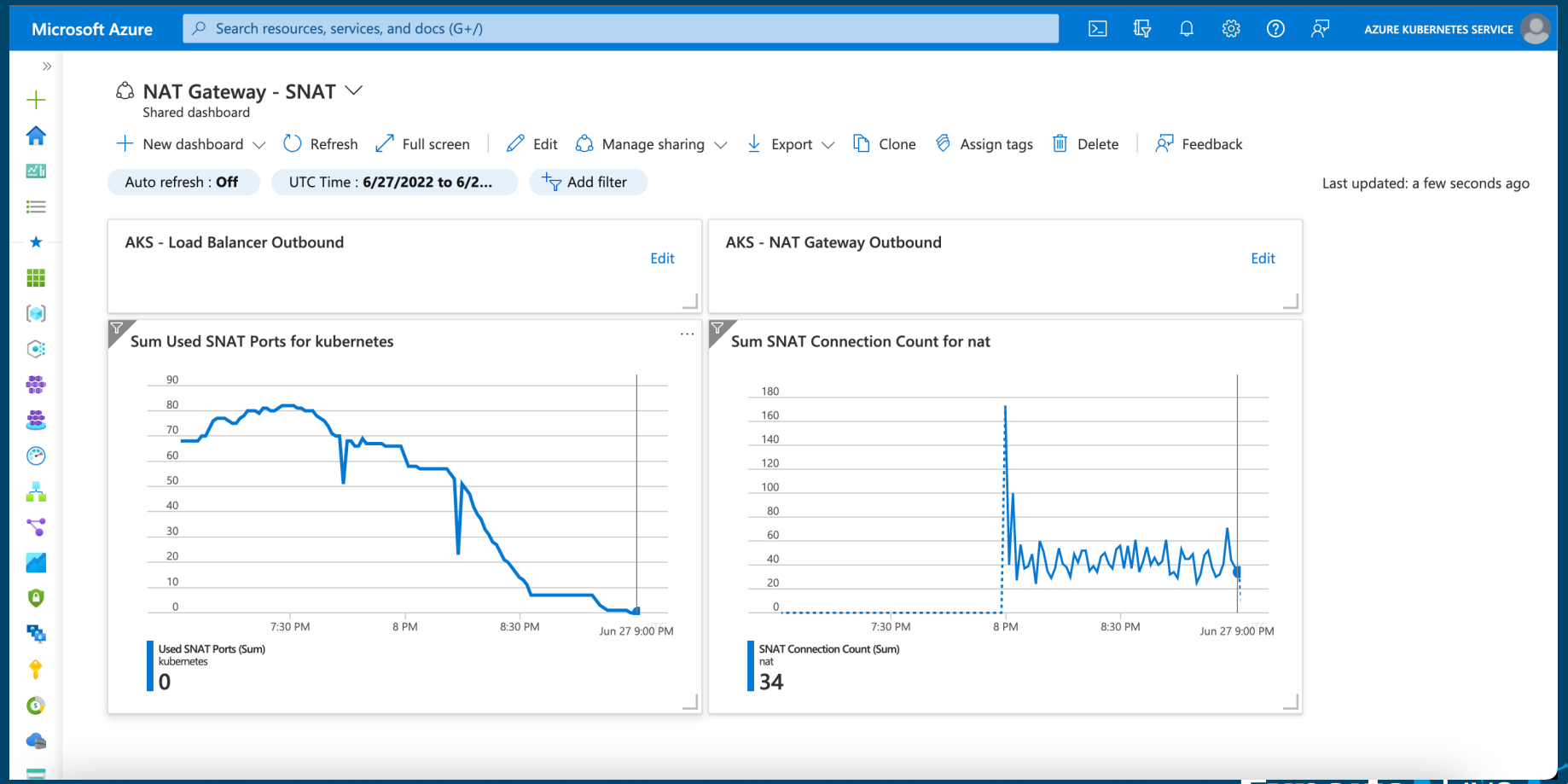
1. Azure Virtual Network NAT hat immer Vorang
   - *"A NAT gateway takes precedence over other outbound scenarios and replaces the default Internet destination of a subnet."*
   - Ausgehender Traffic geht über Virtual Network NAT und nicht über die sogenannten Outbound Rules des Load Balancers

2. Bei *managedNATGateway* oder *userAssignedNATGateway* kann man bei einem Virtual Network NAT Ausfall den AKS Cluster ohne ein Redeployment wiederherstellen
   - Der *outboundType* loadBalancer ermöglicht es Virtual Network NAT vom Subnetz zu entfernen und AKS wird wieder die Outbound Rules des Load Balancers nutzen

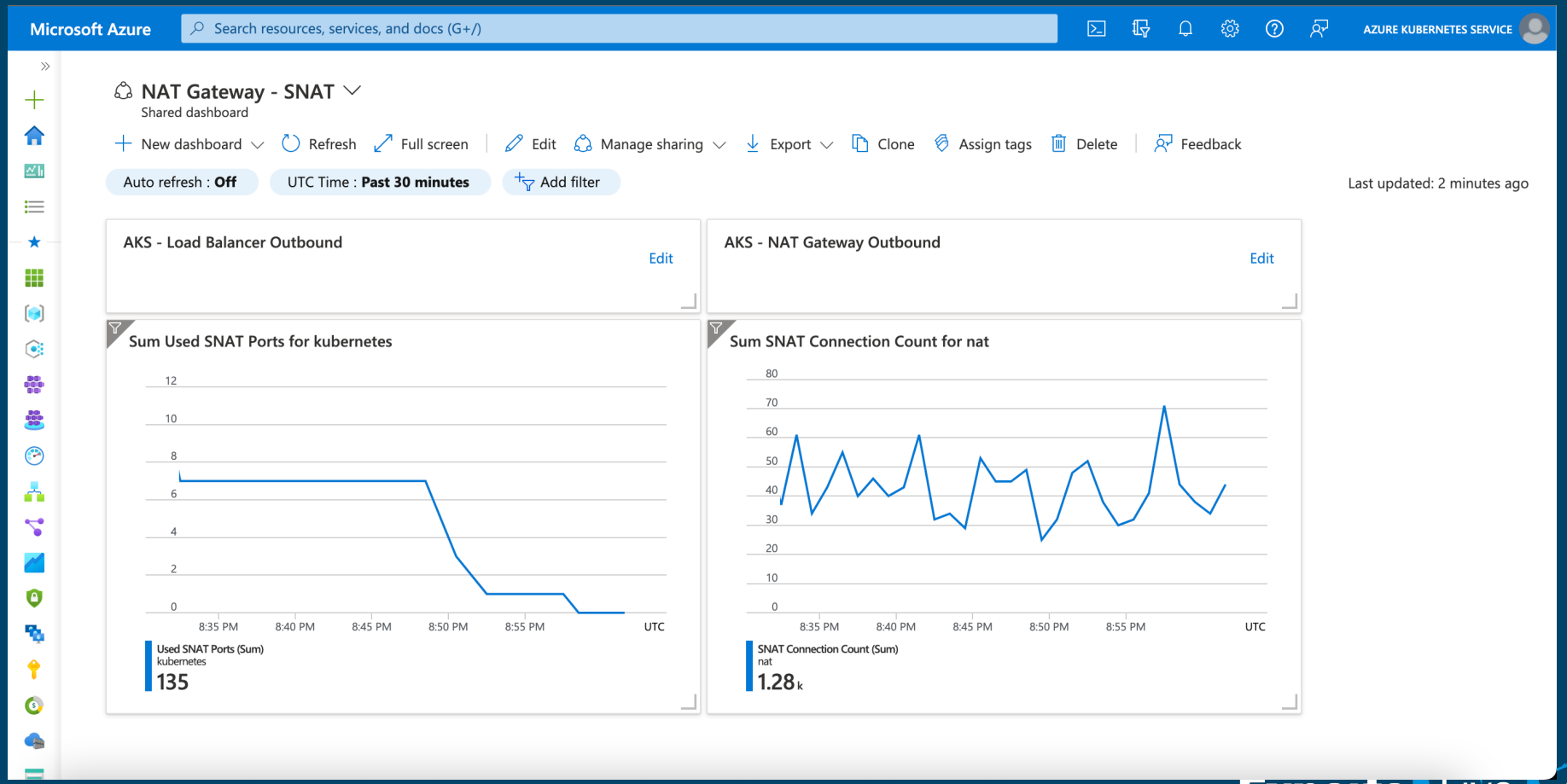
Why should I stick to *outboundType* loadBalancer?



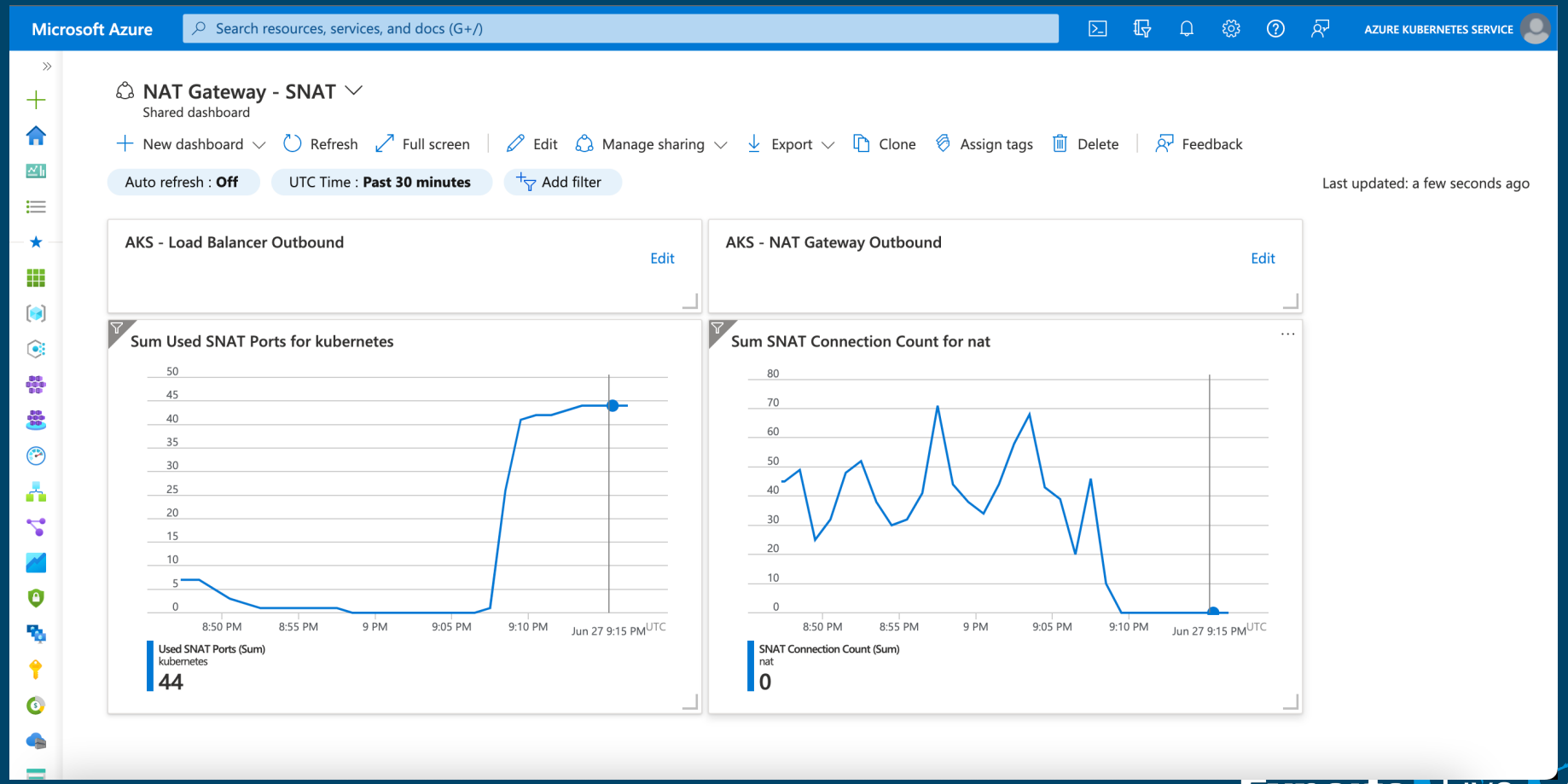
Why should I stick to *outboundType* loadBalancer?



Why should I stick to *outboundType* loadBalancer?



Why should I stick to *outboundType* loadBalancer?



Zusammenfassung

- DR Strategie und PaaS Services
- Developers/Engineers sollten die Kubernetes Prinzipien kennen
- AKS up-to-date halten
- Governance, Monitoring und Alerting

Vielen Dank!

Experts Live Germany

Appendix

- Pod Disruption Budget
  - https://kubernetes.io/docs/concepts/workloads/pods/disruptions/
  - https://www.danielstechblog.io/increase-your-application-availability-with-a-poddisruptionbudget-on-azure-kubernetes-service/
- Pod Topology Spread Constraints
  - https://kubernetes.io/docs/concepts/workloads/pods/pod-topology-spread-constraints/
  - https://www.danielstechblog.io/distribute-your-application-across-different-availability-zones-in-aks-using-pod-topology-spread-constraints/
- Kured
  - https://docs.microsoft.com/en-us/azure/aks/node-updates-kured
- Node image upgrade
  - https://docs.microsoft.com/en-us/azure/aks/node-image-upgrade

Appendix

- Auto-upgrade channel
  - https://docs.microsoft.com/en-us/azure/aks/upgrade-cluster#set-auto-upgrade-channel
- Recommended Alerts
  - https://docs.microsoft.com/en-us/azure/azure-monitor/containers/container-insights-metric-alerts
- SNAT Port Exhaustion
  - https://docs.microsoft.com/en-us/azure/load-balancer/load-balancer-outbound-connections
  - https://docs.microsoft.com/en-us/azure/load-balancer/load-balancer-standard-diagnostics#how-do-i-check-my-snat-port-usage-and-allocation
  - https://www.danielstechblog.io/detecting-snat-port-exhaustion-on-azure-kubernetes-service/
- Uptime SLA
  - https://docs.microsoft.com/en-us/azure/aks/uptime-sla