AKS Hero Updates

Kunal Chandratre (Sr. Cloud Solution Architect)
@Microsoft

Microsoft

Disclaimer*

- All views expressed in this session are personal, in no way it represents the company I work for.

Agenda

- Orchestration in containers
- AKS Updates
- AKS and AI
- Demos
- Wrap up!

Is container a new concept?

Popular Orchestrators

Docker Swarm

Apache Mesos – Marathon

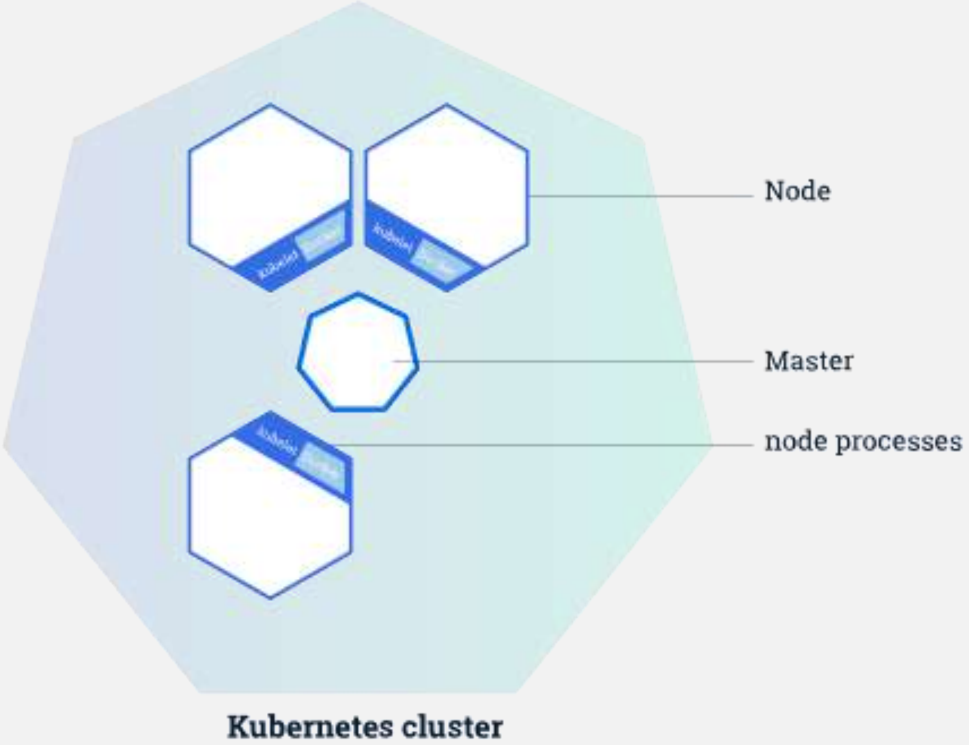
Kubernetes

Microsoft

What is Kubernetes?

"Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications."

Kubernetes comes from the Greek word **κυβερνήτης:**, which means *helmsman* or *ship pilot*, ie: the captainer of a container ship.

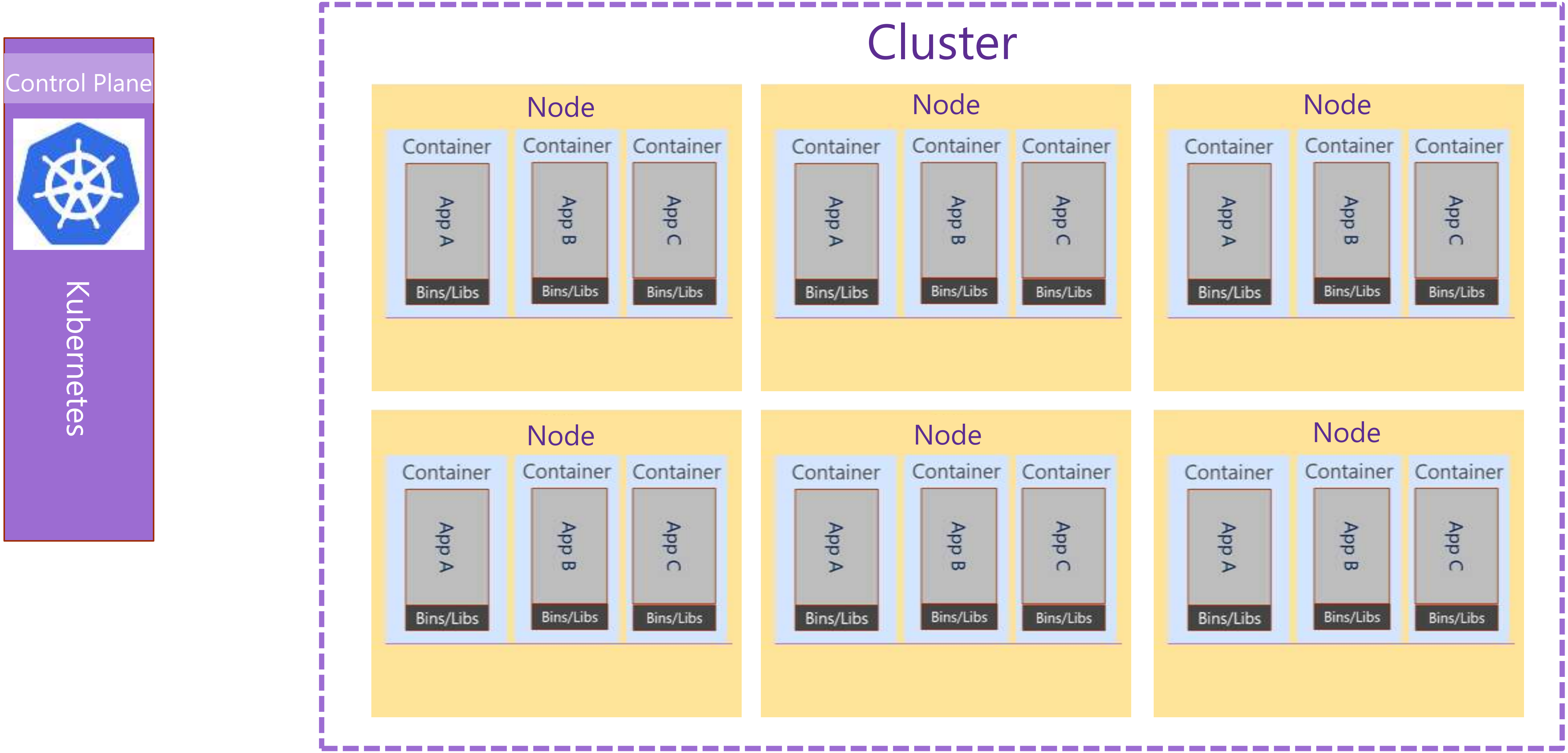


Kubernetes

- **Master**: The server that runs the Kubernetes management processes, including the API service, replication controller and scheduler.
- **Node**: The host that runs the kubelet service and the Docker Engine. Minions receive commands from the master.
- **Kubelet**: The node-level manager in Kubernetes; it runs on a minion.
- **Pod**: The collection of containers deployed on the same minion.
- **Replication controller**: Defines the number of pods or containers that need to be running.
- **Service**: A definition that allows the discovery of services/ports published by each container, along with the external proxy used for communications.
- **Kubecfg**: The command line interface that talks to the master to manage a Kubernetes deployment.

Note – In any orchestrator High Availability of Infra has to be managed by you.

AKS is Kubernetes (Control Plane) as a Service



Control Plane

Kubernetes

Cluster

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

Node

Container | Container | Container
App A | App B | App C
Bins/Libs | Bins/Libs | Bins/Libs

AKS Hero Features - Updates

AKS Policies

- https://learn.microsoft.com/en-us/azure/aks/policy-reference

- Security baseline - https://learn.microsoft.com/en-us/security/benchmark/azure/baselines/aks-security-baseline

Example - AKS Policy to save cost

1. Assign policy CPU Memory resource limits for PODs.
2. Deploying pods with higher cpu/ memory threshold than set limit, should fail.
3. Deploying pods with equal to or below cpu/ memory threshold than set limit, should succeed.

Cluster Optimization workbook for Azure Monitor Container Insights

Azure Monitor Container insights for collection of logs and events now has capability to help you optimize your cluster

Detect liveness probe failures

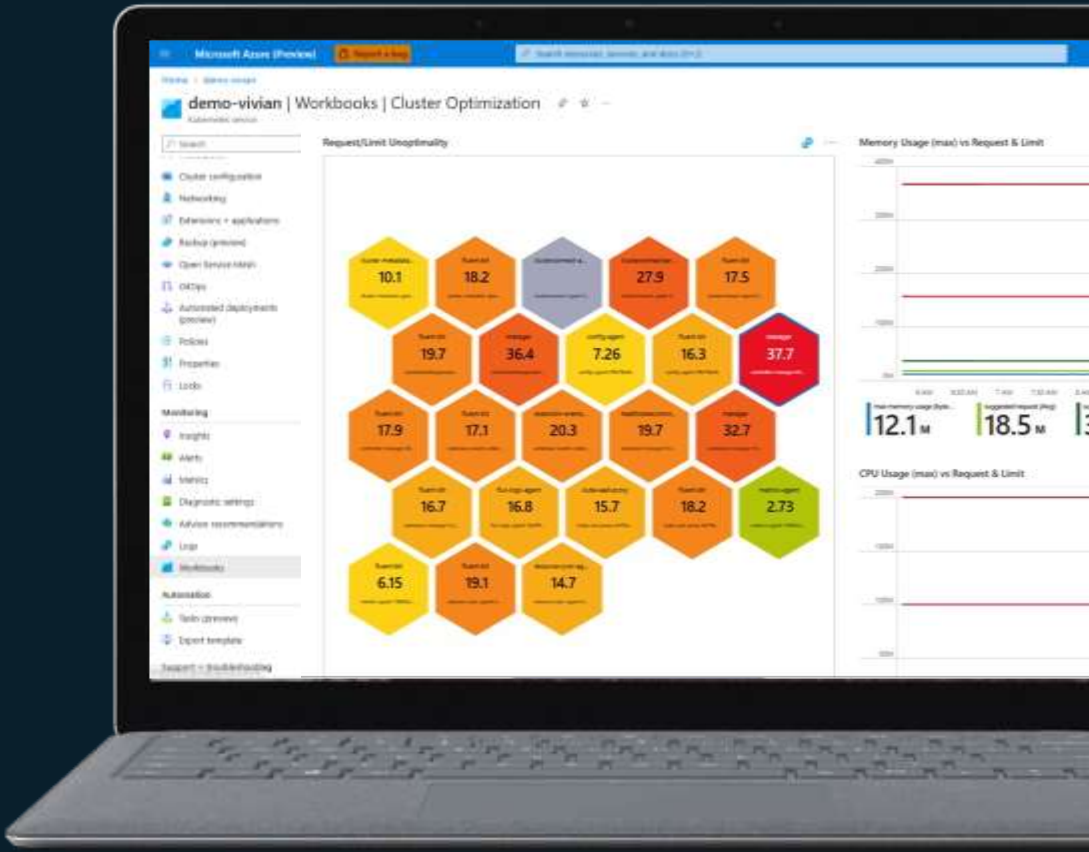
Detect liveness probe failures and their frequencies.

Identify event anomalies

Identify and group event anomalies that indicate recent increases in event volume for easier analysis.

Optimize container limits and requests

Identify containers with high or low CPU and memory limits and requests, along with suggested limit and request values.

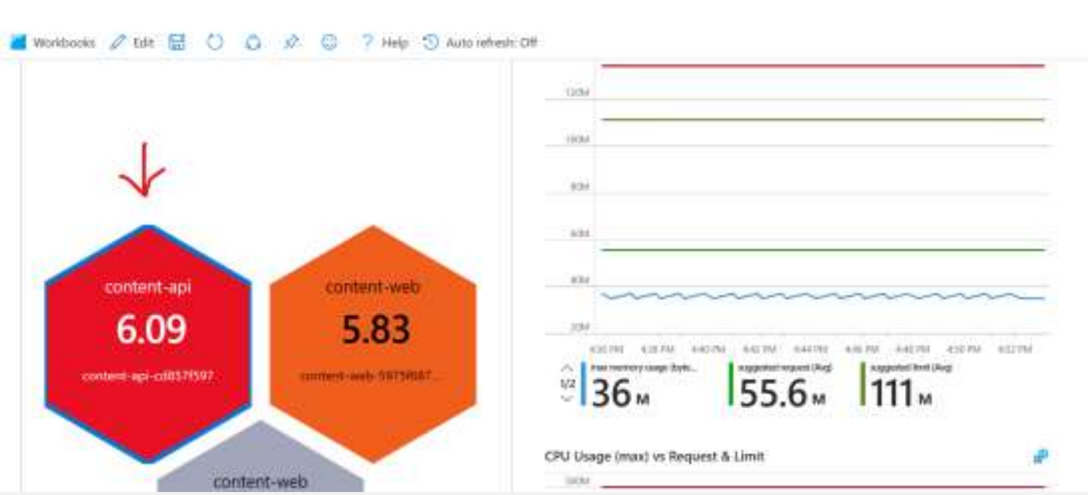


demo-vivian | Workbooks | Cluster Optimization

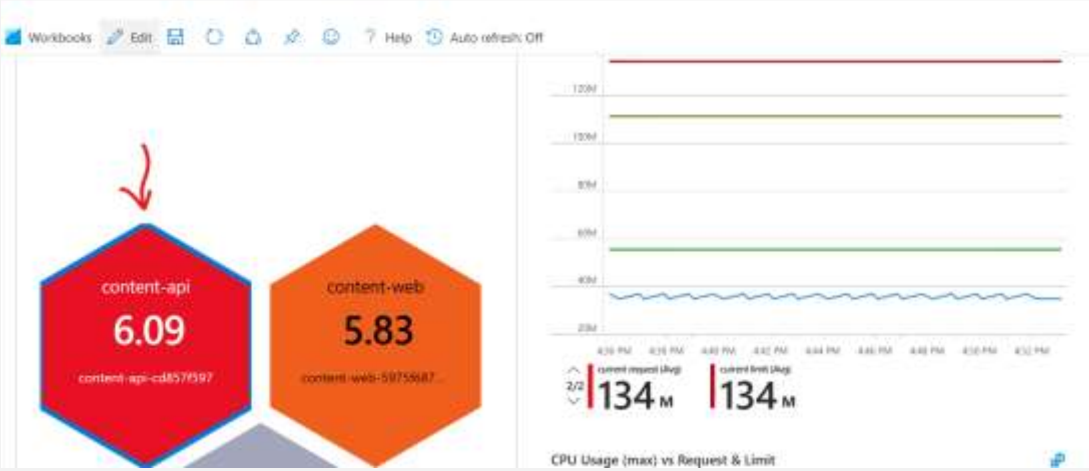
Container Optimizer Example

For this container – Memory -

Max used – 36MB
Suggested request – 55MB
Suggested limit – 111MB
Current request – 134MB
Current limit – 134M



CPU Usage (max) vs Request & Limit



CPU Usage (max) vs Request & Limit

Demo – Cluster Optimization Workbook

1. Red – excessive request and limit is assigned to containers within a pod
2. Green – well set request and limit
3. Gray – no limit or request is set
4. Value closure to zero would be better
5. Show on portal

AKS Cost Analysis

aka.ms/aks/cost-analysis

Azure native experience for cost visibility and allocation

Built on top of open source, vendor neutral CNCF sandbox project OpenCost

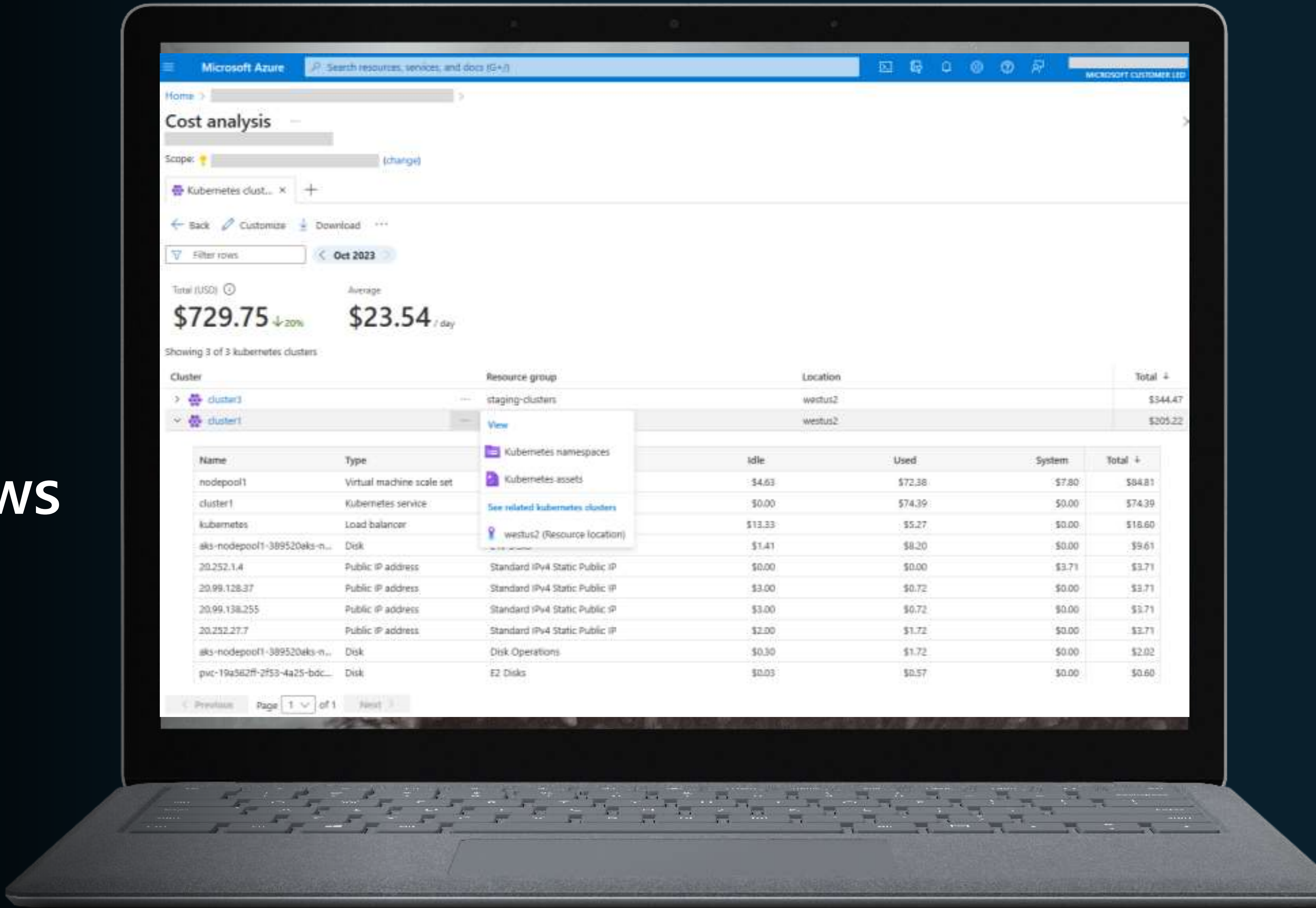
Available for Standard and Premium tier AKS clusters at no additional cost

Ensure costs are allocated to the right teams to drive accountability

Identify high spend areas and opportunities to optimize costs

Proactively identify cost anomalies to prevent unanticipated overspending

Kubernetes specific views



Home > [redacted] >

Cost analysis

[redacted]

Scope: [redacted] (change)

Total (USD) $729.75 ↓20%

Average $23.54 / day

Showing 3 of 3 kubernetes clusters

Cluster	Resource group	Location	Total ↓
cluster3	staging-clusters	westus2	$344.47
cluster1		westus2	$205.22

Name	Type		Idle	Used	System	Total ↓
nodepool1	Virtual machine scale set		$4.63	$72.38	$7.80	$84.81
cluster1	Kubernetes service		$0.00	$74.39	$0.00	$74.39
kubernetes	Load balancer		$13.33	$5.27	$0.00	$18.60
aks-nodepool1-389520aks-n...	Disk	[illegible]	$1.41	$8.20	$0.00	$9.61
20.252.1.4	Public IP address	Standard IPv4 Static Public IP	$0.00	$0.00	$3.71	$3.71
20.99.128.37	Public IP address	Standard IPv4 Static Public IP	$3.00	$0.72	$0.00	$3.71
20.99.138.255	Public IP address	Standard IPv4 Static Public IP	$3.00	$0.72	$0.00	$3.71
20.252.27.7	Public IP address	Standard IPv4 Static Public IP	$2.00	$1.72	$0.00	$3.71
aks-nodepool1-389520aks-n...	Disk	Disk Operations	$0.30	$1.72	$0.00	$2.02
pvc-19a562ff-2f53-4a25-bdc...	Disk	E2 Disks	$0.03	$0.57	$0.00	$0.60

Kubernetes cluster view

Demo – Aks Cost Analysis

1. Show on portal.

Azure Arc Enabled Kubernetes

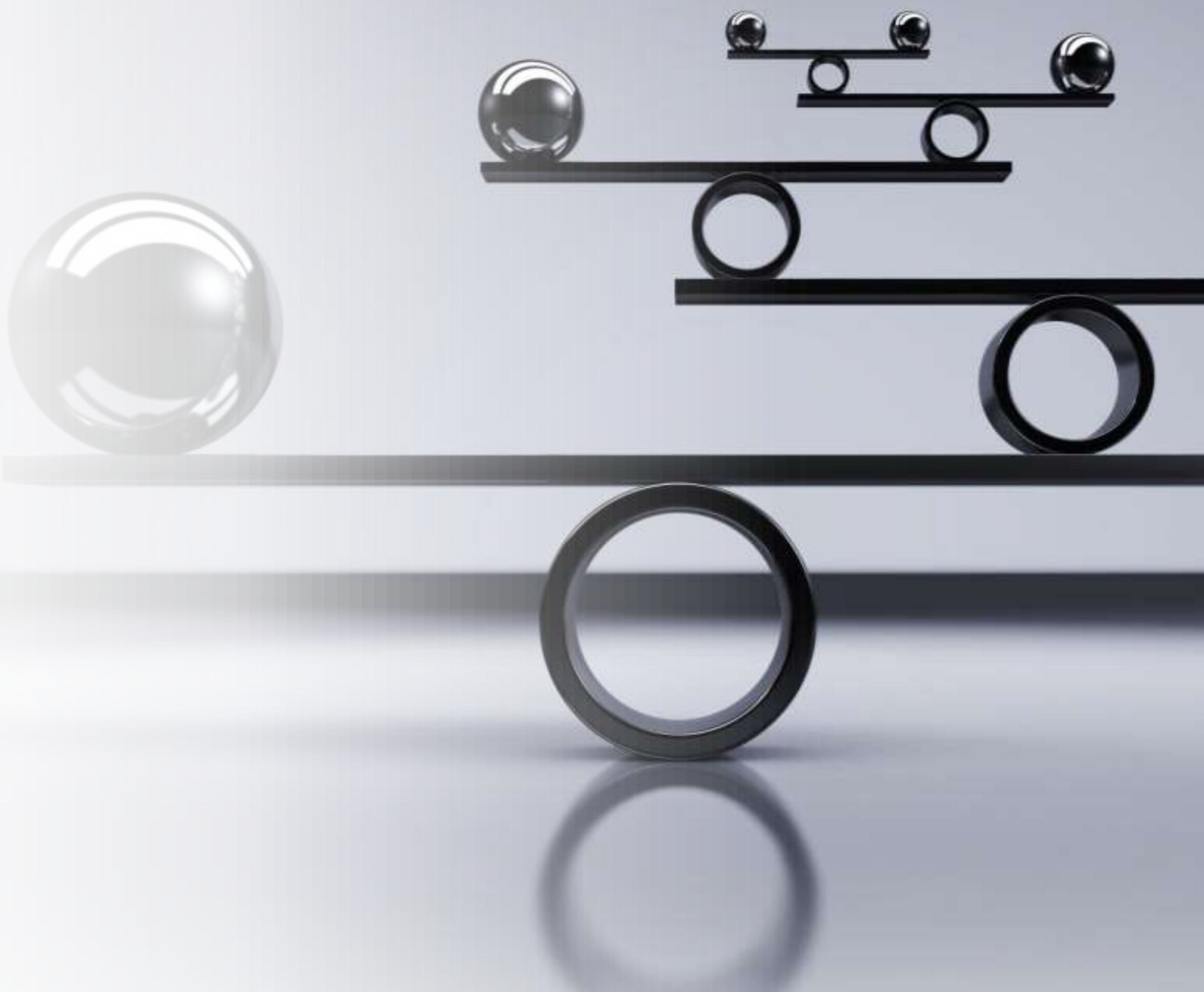
- Connect any Kubernetes deployment to Azure
- Traffic flows over https – private endpoint is in preview
- Enables below features –
  - Defender
  - Policies
  - AAD
  - Cluster Connect – Connect to cluster without opening inbound ports
  - Kubernetes Partners distribution supported - https://learn.microsoft.com/en-us/azure/azure-arc/kubernetes/validation-program

AKS Construction Helper

- https://azure.github.io/AKS-Construction/?default=es
- https://github.com/Azure/Aks-Construction

There is difference in infra scale and App based infra scale

KEDA - https://keda.sh/docs/2.8/scalers/



Azure Linux

Container Host OS for Azure Kubernetes Service

Just enough OS

Smaller
OS footprint

Reliability

Shift left in build pipeline

Prevent defective builds from advancing

Stringent package tests

Performance tests

Security

Secure defaults

Fast CVE patching

Secure supply chain

Upcoming

In-place migration to Azure Linux

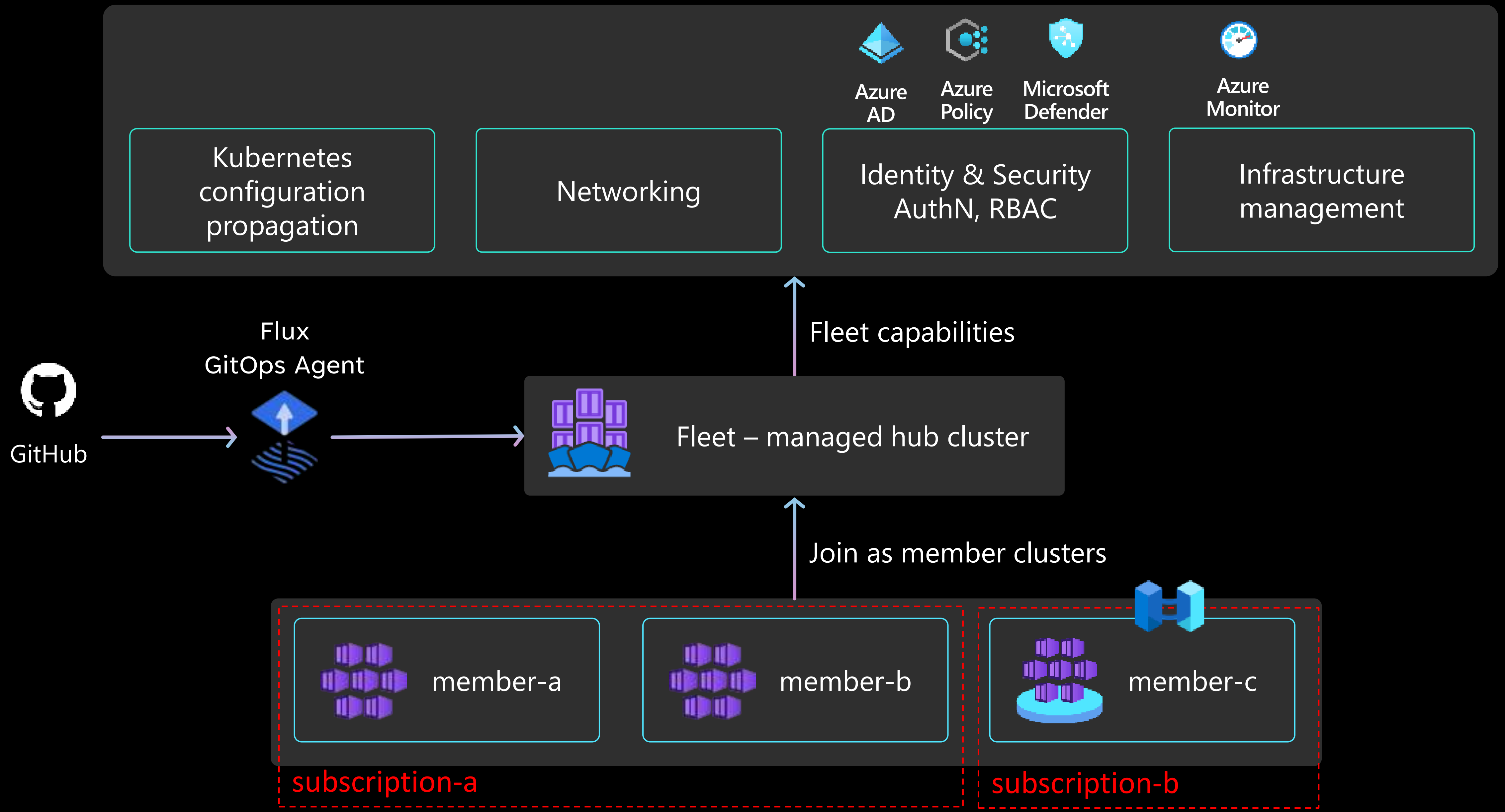
Working in collaboration with the Azure Linux team, we’ll be releasing a feature to make it easier for you to migrate your workloads from Ubuntu to Azure Linux on an existing nodepool.

aka.ms/aks/ossku-in-place-migration

Fleet Manager

- Why?
  - Managing 20+ AKS clusters in any org is always pain
  - Complexity increases in case of multi cloud, on premises Kubernetes
  - No Central governance exists as of today across all clusters across subscriptions, regions, resource groups etc.
- Join cluster across RG, Sub, Regions
- Member clusters should be under same AAD tenant.
- Support for multi cloud, hybrid AKS clusters
- Selective member configuration supported
- Max 20 member as of today
- Sample Use cases –
  - Upgrade AKS version of all clusters
  - Create same namespace across all clusters. Example, same namespace for ingress
  - Create common RBAC across all clusters
  - Centralised pod to pod communication policies

Azure Kubernetes Fleet Manager



AKS Kubernetes version Long Term Support (LTS)

Two years of Microsoft support, including CVEs and critical bugs

Upgrade available to the next AKS Kubernetes LTS

Ability to return to the upstream version train

Forked after upstream EOL, maintained by Microsoft in the open

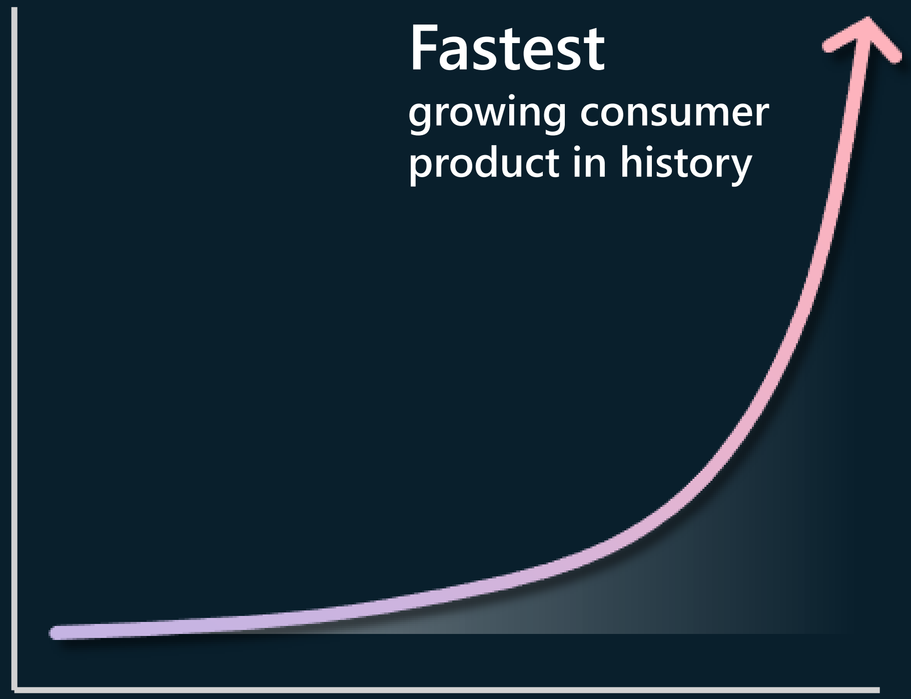
Azure powers OpenAI and ChatGPT

ChatGPT

Runs on Azure Kubernetes Service (AKS)

Backed by Azure Cosmos DB

Developed on GitHub



Fastest
growing consumer
product in history

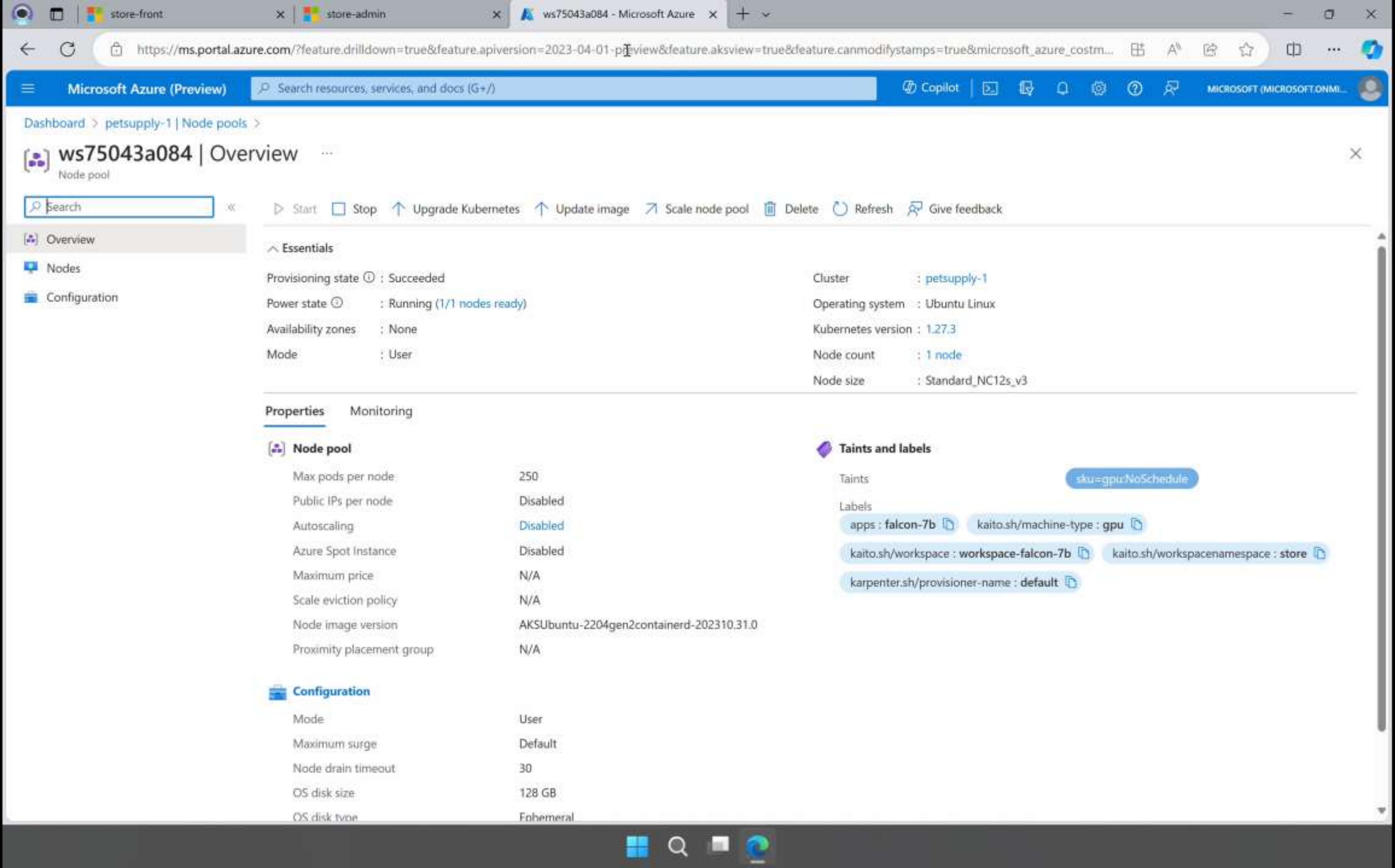
Every app will be reinvented with AI

New apps will be built that weren’t possible before

Microsoft Copilot for Azure

An AI companion that simplifies how you design, operate, optimize, and troubleshoot both apps and infrastructure from cloud to edge.

Available initially in Azure portal. Expanding to Azure mobile app and CLI.


store-front
store-admin
ws75043a084 - Microsoft Azure
https://ms.portal.azure.com/?feature.drilldown=true&feature.apiversion=2023-04-01-preview&feature.aksview=true&feature.canmodifystamps=true&microsoft_azure_costm...
Microsoft Azure (Preview)
Search resources, services, and docs (G+/)
Copilot
MICROSOFT (MICROSOFT.ONMI...
Dashboard > petsupply-1 | Node pools >
Search
Overview
Nodes
Configuration
Start
Stop
Upgrade Kubernetes
Update image
Scale node pool
Delete
Refresh
Give feedback
Essentials
Properties
Monitoring

ws75043a084 | Overview

Node pool

Provisioning state	Succeeded
Power state	Running (1/1 nodes ready)
Availability zones	None
Mode	User
Cluster	petsupply-1
Operating system	Ubuntu Linux
Kubernetes version	1.27.3
Node count	1 node
Node size	Standard_NC12s_v3

Node pool

Max pods per node	250
Public IPs per node	Disabled
Autoscaling	Disabled
Azure Spot Instance	Disabled
Maximum price	N/A
Scale eviction policy	N/A
Node image version	AKSUbuntu-2204gen2containerd-202310.31.0
Proximity placement group	N/A

Taints and labels

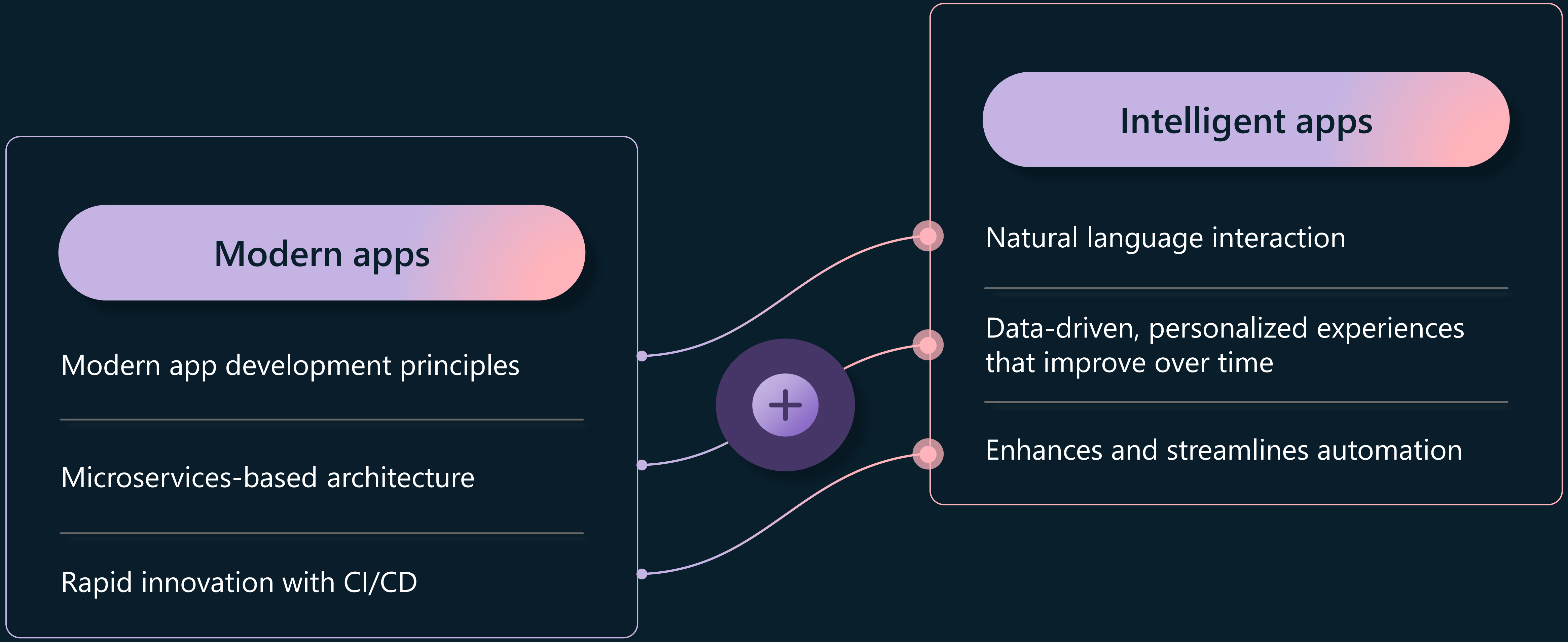
Taints: sku=gpu:NoSchedule

Labels:
- apps : falcon-7b
- kaito.sh/machine-type : gpu
- kaito.sh/workspace : workspace-falcon-7b
- kaito.sh/workspacenamespace : store
- karpenter.sh/provisioner-name : default

Configuration

Mode	User
Maximum surge	Default
Node drain timeout	30
OS disk size	128 GB
OS disk type	Ephemeral

Generative AI makes apps truly intelligent



Modern apps

- Modern app development principles
- Microservices-based architecture
- Rapid innovation with CI/CD

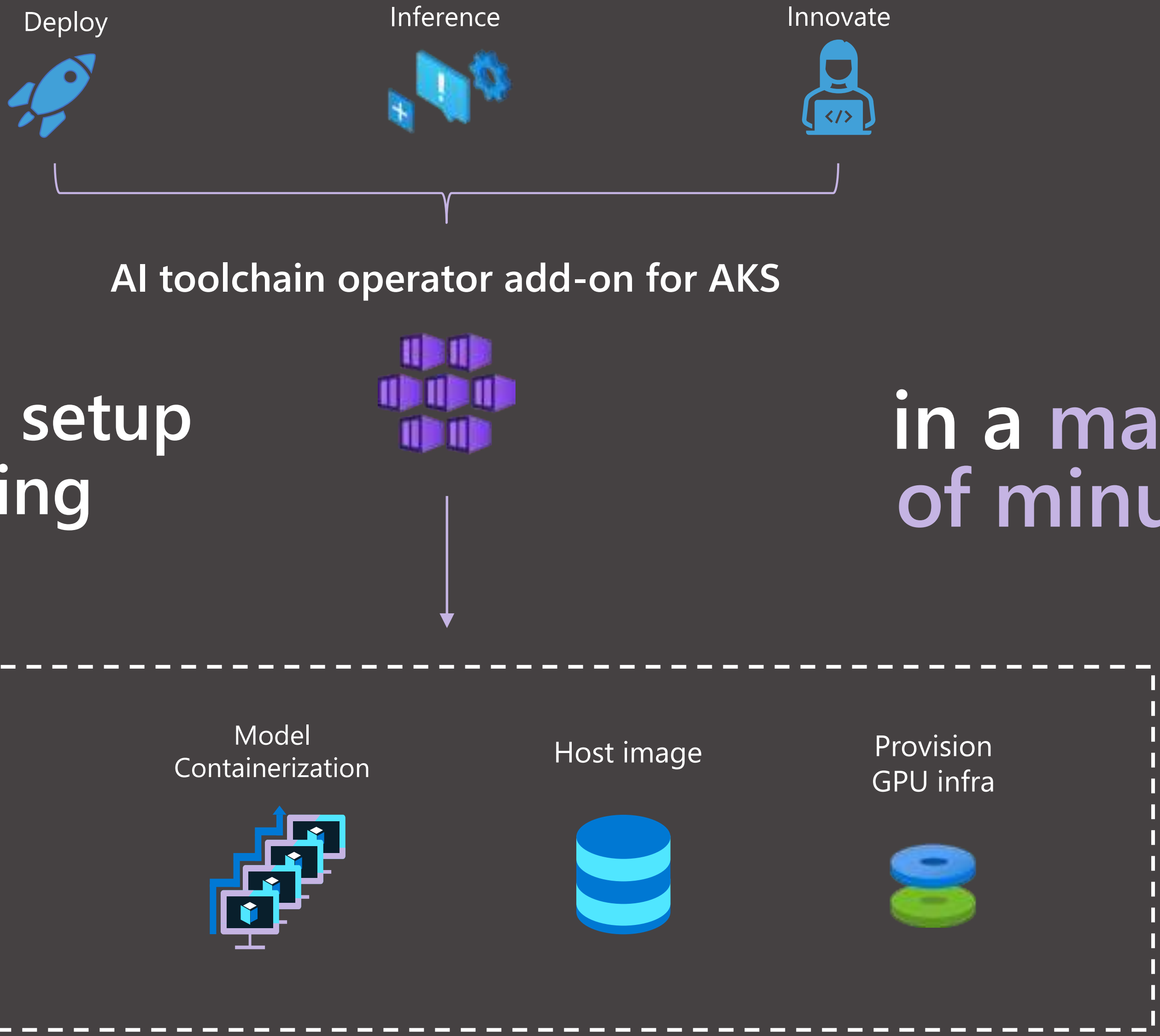
Intelligent apps

- Natural language interaction
- Data-driven, personalized experiences that improve over time
- Enhances and streamlines automation

AI toolchain operator add-on for AKS

Announcing

Workspace and infra setup to model inferencing



in a matter of minutes

Thank you…

Stay connected…

http://sanganakauthority.blogspot.com/

Twitter - @KunalChandratre

Active on LinkedIn!!