Microsoft Azure Aministrator: Configure Virtual Machines for High Availability and Scalability

CONFIGURE HIGH AVAILABILITY AND SCALABILITY

Michael Teske
AUTHOR EVANGELIST-CLOUD ENGINEER, PLURALSIGHT
@teskemj

Course Coverage of Certification Objectives

Configure High Availability and Scability

- Configure VMs for high availability
  - Availability sets and zones
  - Fault and update domains
- Deploy and configure scale sets

Full certification exam skills outline available at http://bit.ly/az104ms

Exercise Files

Slides

Code

Links to Resources

Configure for High Availability

High Availability Constructs

Availability Zones

Fault Domains

Update Domains

Availability Sets

Scale Sets

Availability Zones

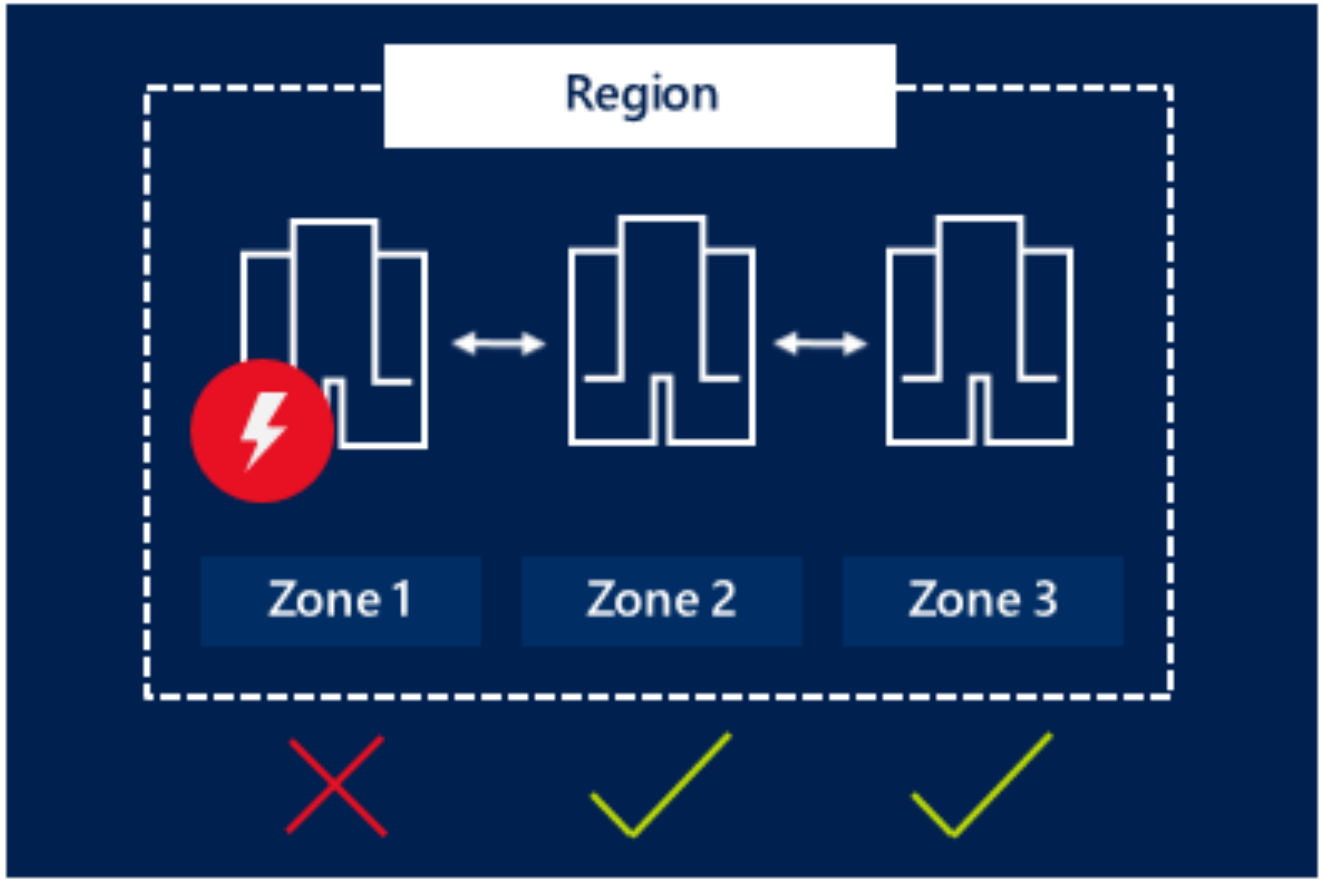
Availability zones distribute VMs across *Azure regions*

- Good for application deployments
- 3 zones per region

Standard SKU load balancers are availability zone aware

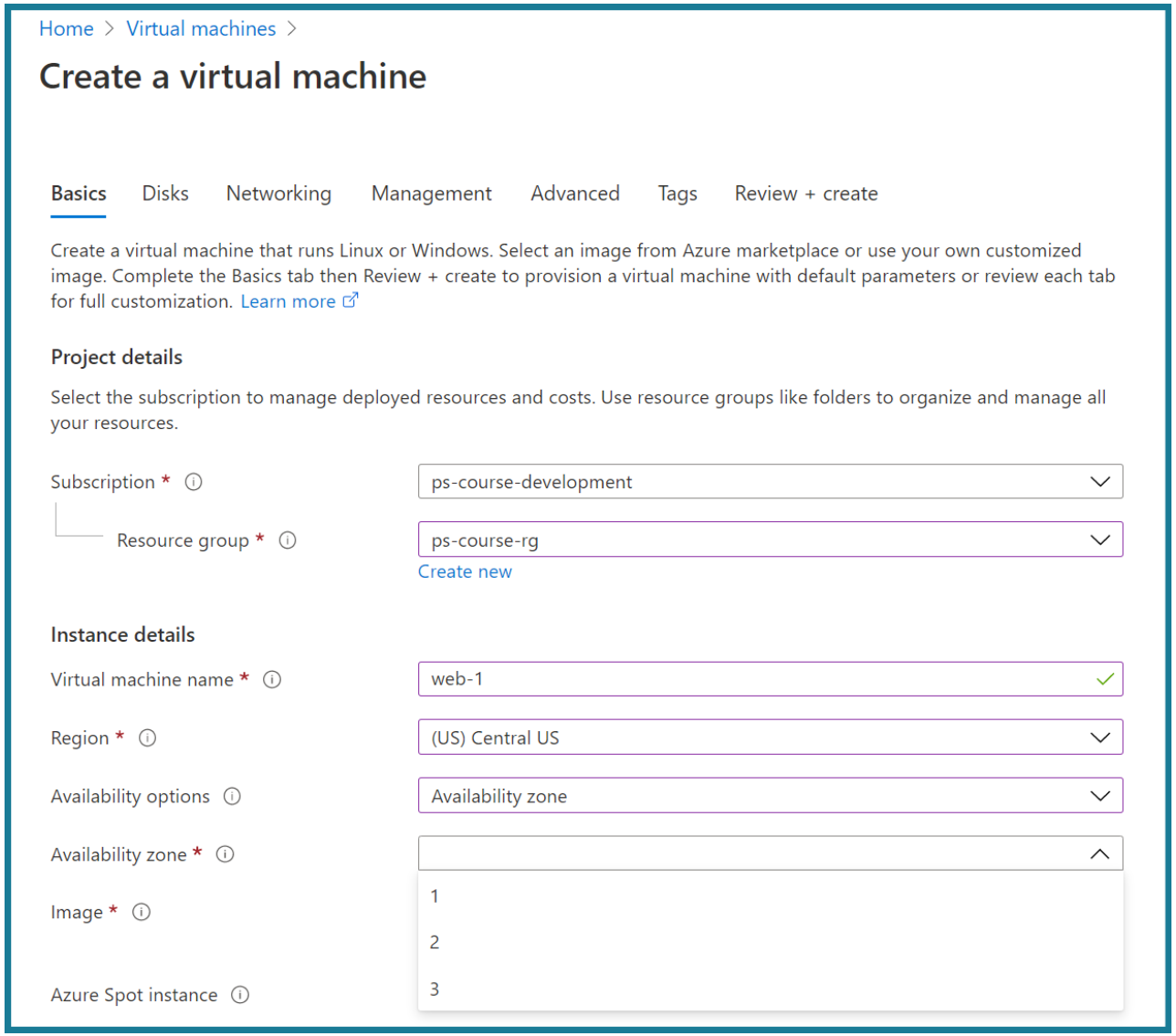
Standard SKU PIPs are required

Availability Zones



Reference: https://bit.ly/2Izf9s9

Availability Zones



SLA Availability Zones

"For all Virtual Machines that have two or more instances deployed across two or more Availability Zones in the same Azure region, we guarantee you will have Virtual Machine Connectivity to at least one instance at least 99.99% of the time."

Reference: https://bit.ly/2IrIG6S

Fault Domains

Logical group of hardware in an Azure datacenter

VMs in the same fault domain share common power source and physical network switch

Update Domains

Update domains

- Protect against normal maintenance updates
  - Applying updates to hardware

VMs created in the same update domain will be restarted together during planned maintenance

Only one update domain restarted at a time

Availability Sets

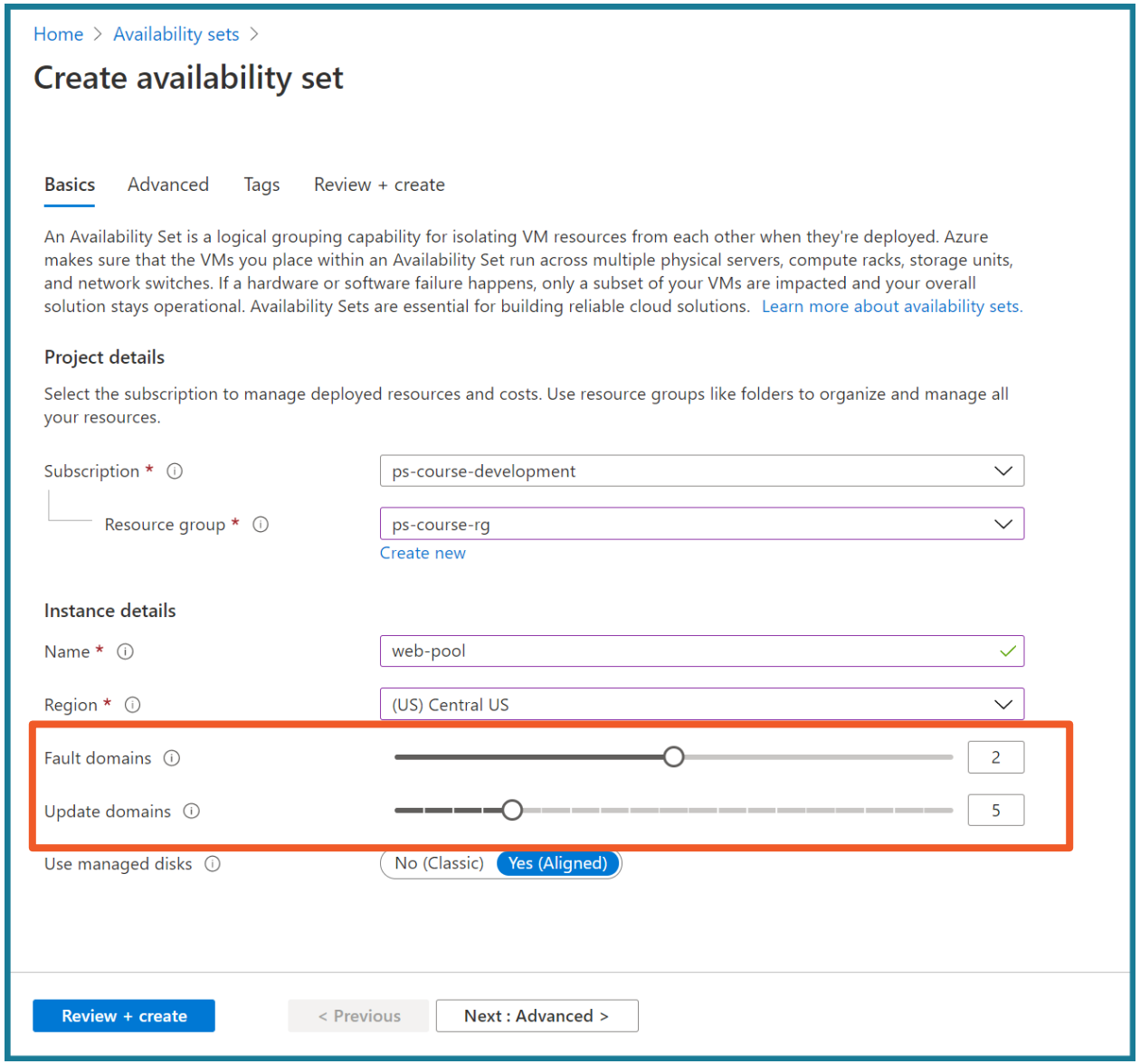
Availability sets group VMs to distribute across a single datacenter

- Minimize disruptions caused by maintenance or outages
- Good for VMs

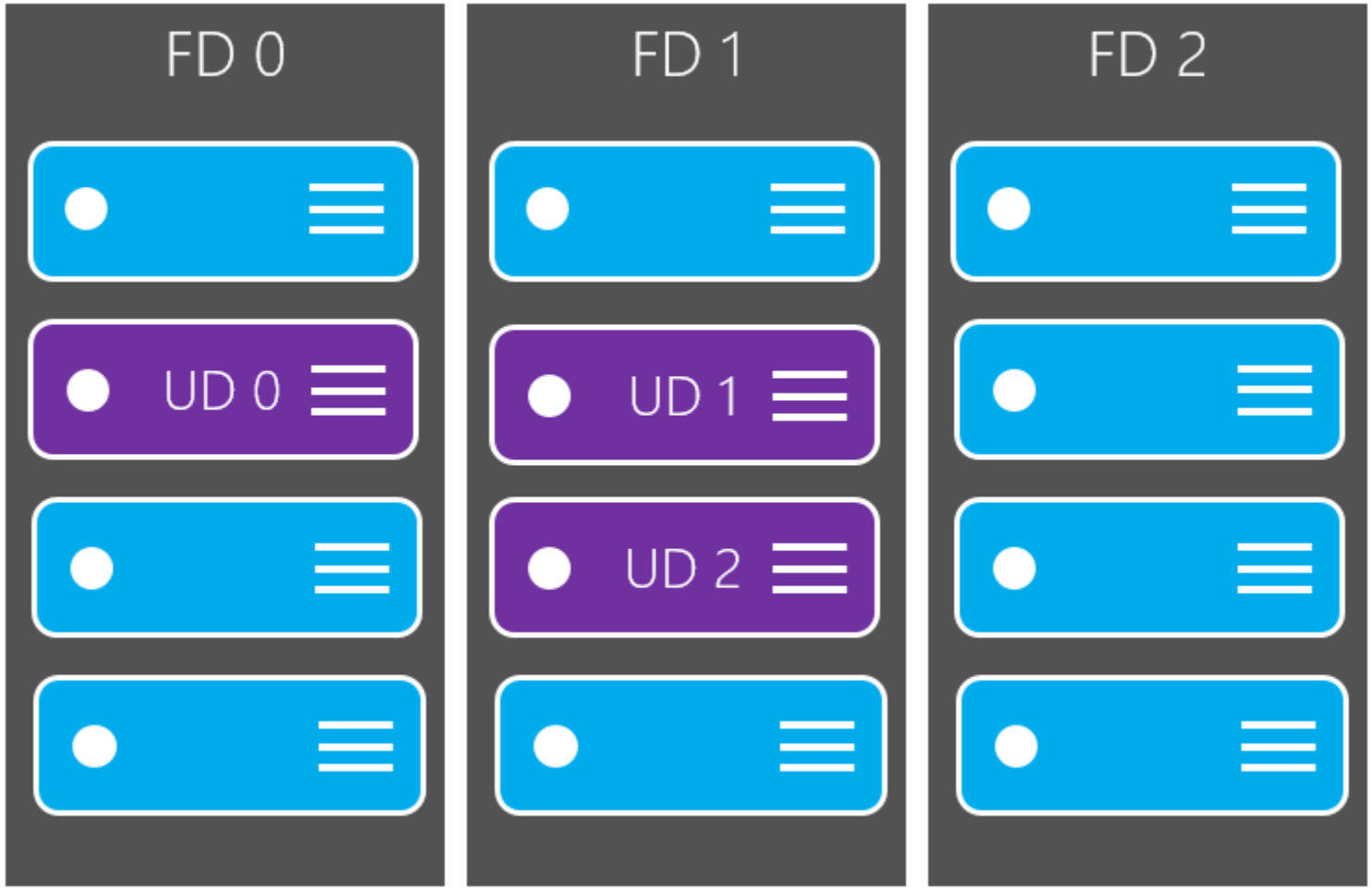
Cannot add a VM to availability set post deployment.

- Must be done at creation

Availability Set

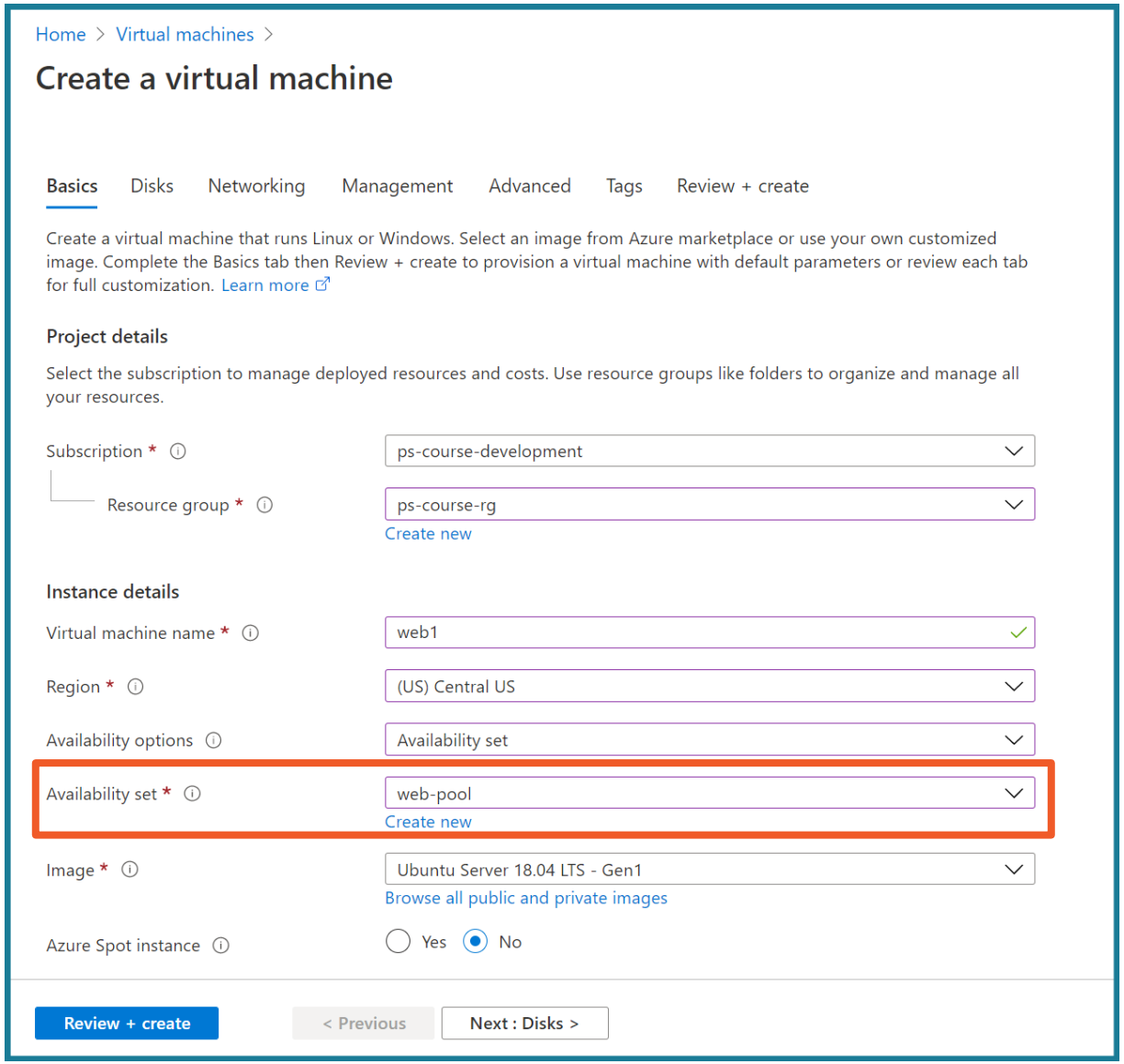


Availability Sets



Reference: https://bit.ly/2Izf9s9

Availability Sets



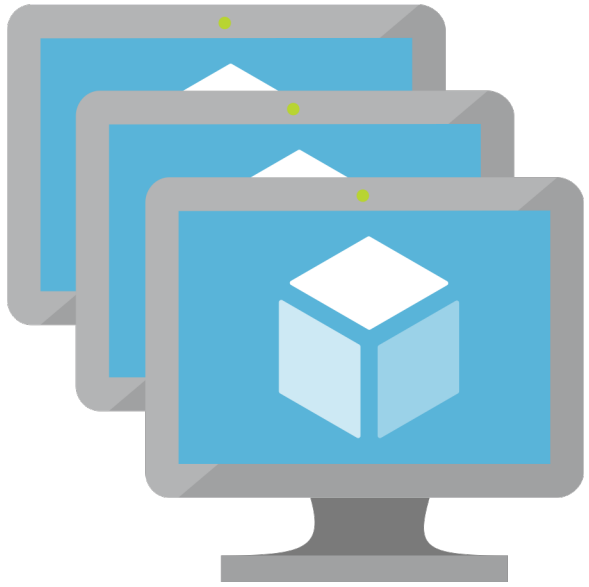
SLA for Availability Sets

For all Virtual Machines that have two or more instances deployed in the same Availability Set or in the same Dedicated Host Group, we guarantee you will have Virtual Machine Connectivity to at least one instance at least 99.95% of the time.

Reference: https://bit.ly/2IrIG6S

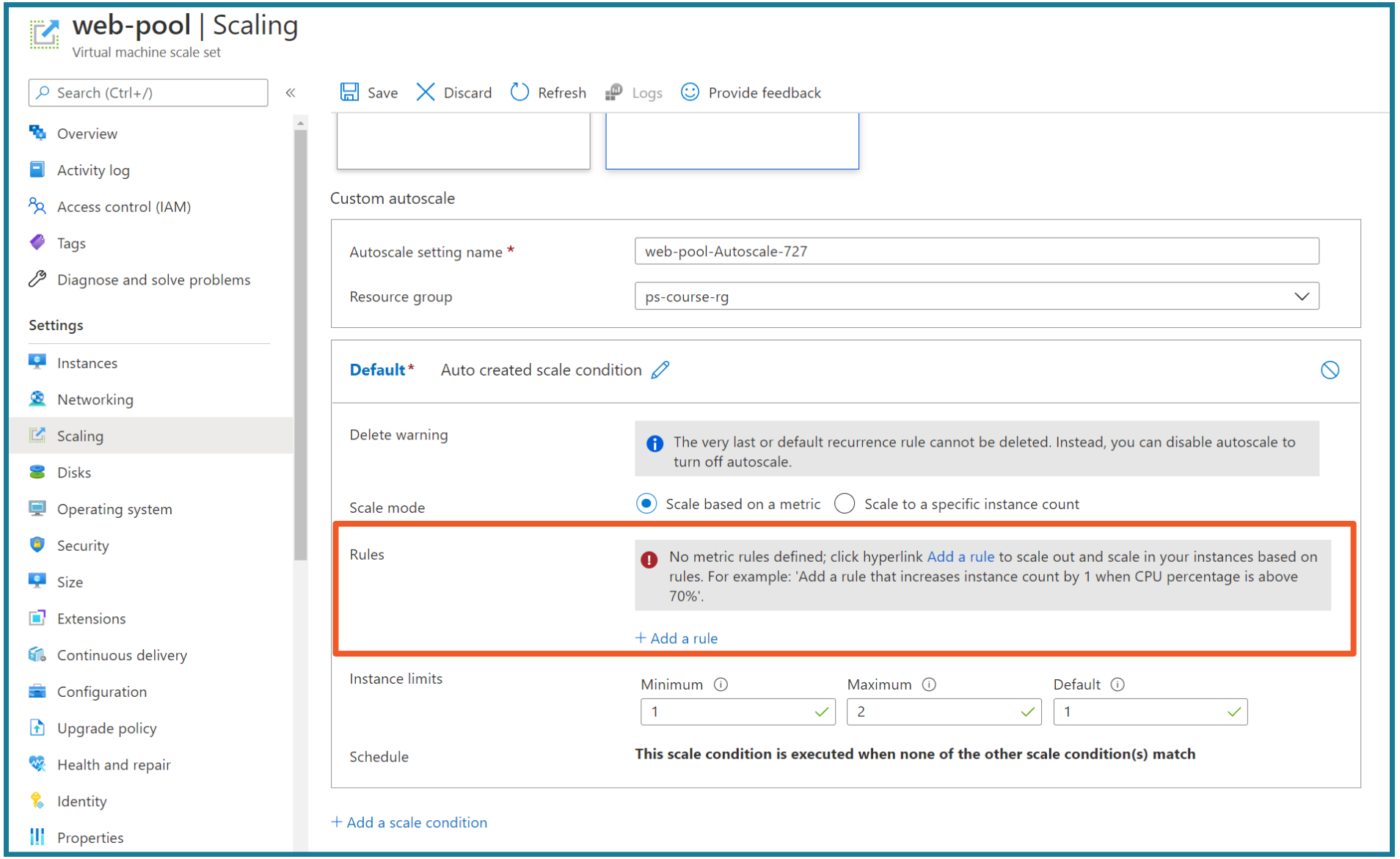
Configure for High Scalability

Virtual Machine Scale Sets

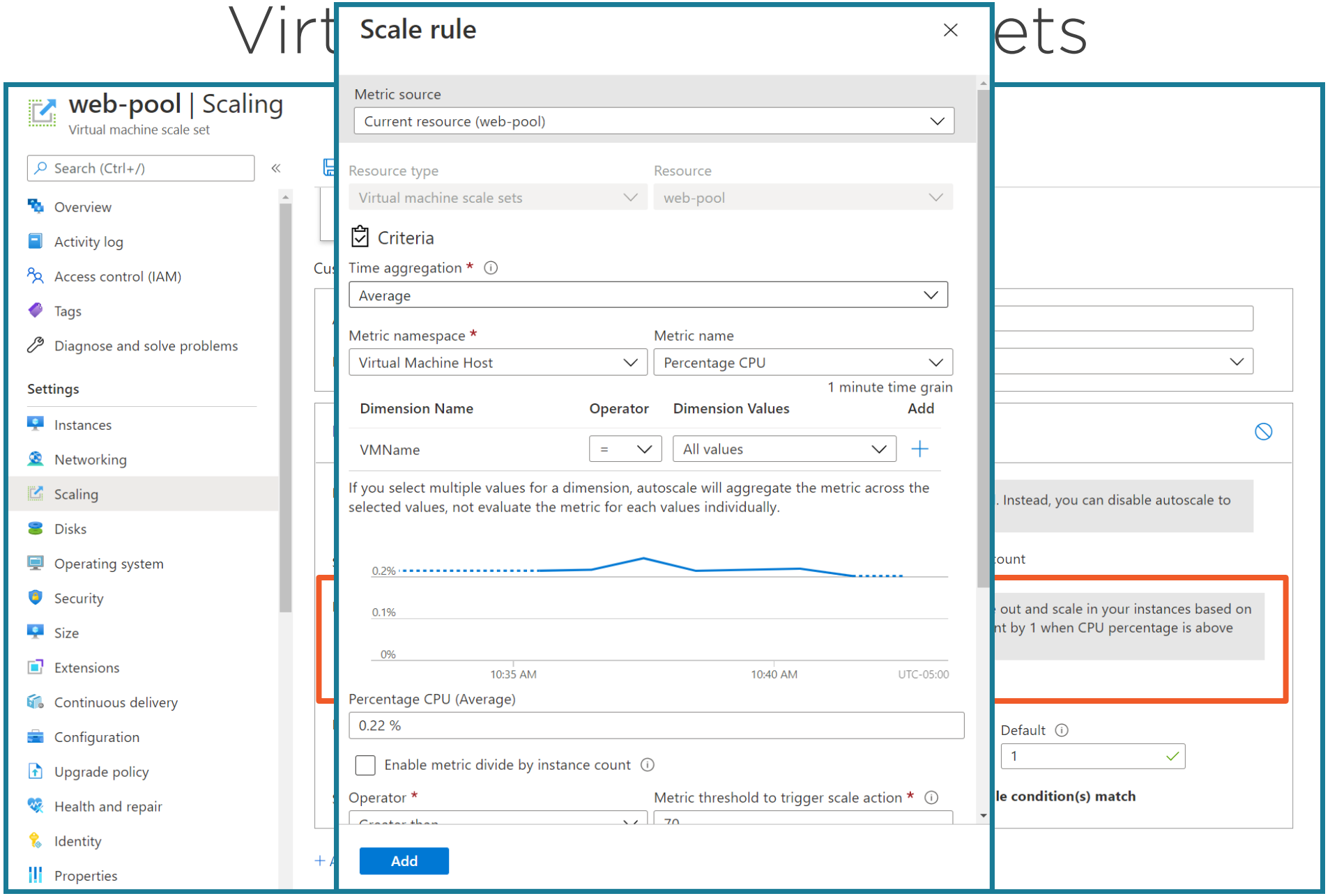


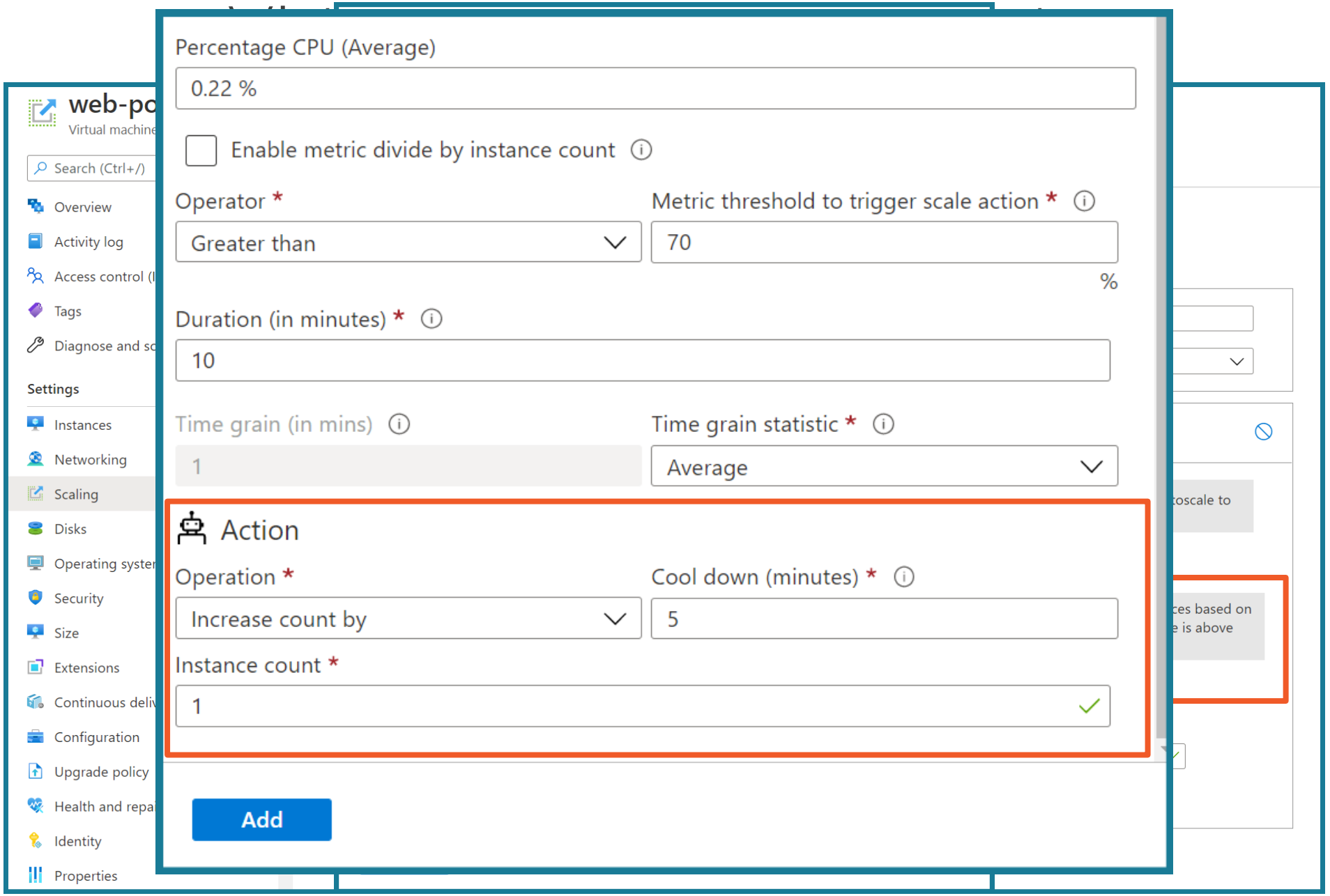
- Group of load balanced virtual machines
- Can scale automatically based on demand or schedule
- 2 or more VMs recommended
- No additional cost other than extra instances
- Can be deployed across multiple update/fault domains

Virtual Machine Scale Sets

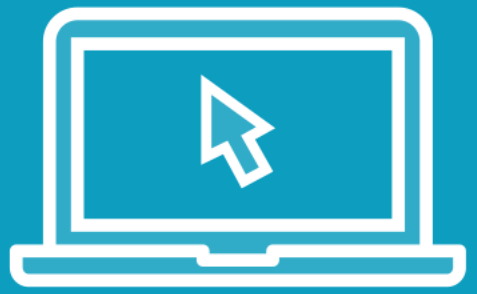


Virt ets


web-pool | Scaling
Virtual machine scale set
Search (Ctrl+/)
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Instances
Networking
Scaling
Disks
Operating system
Security
Size
Extensions
Continuous delivery
Configuration
Upgrade policy
Health and repair
Identity
Properties
Scale rule
Metric source
Current resource (web-pool)
Resource type
Virtual machine scale sets
Resource
web-pool
Criteria
Time aggregation *
Average
Metric namespace *
Virtual Machine Host
Metric name
Percentage CPU
1 minute time grain
Dimension Name
Operator
Dimension Values
Add
VMName
=
All values
If you select multiple values for a dimension, autoscale will aggregate the metric across the selected values, not evaluate the metric for each values individually.
0.2%
0.1%
0%
10:35 AM
10:40 AM
UTC-05:00
Percentage CPU (Average)
0.22 %
Enable metric divide by instance count
Operator *
Metric threshold to trigger scale action *
Add
. Instead, you can disable autoscale to
count
out and scale in your instances based on
nt by 1 when CPU percentage is above
Default
1
le condition(s) match


web-po
Virtual machine
Search (Ctrl+/)
Overview
Activity log
Access control (I
Tags
Diagnose and so
Settings
Instances
Networking
Scaling
Disks
Operating syster
Security
Size
Extensions
Continuous deliv
Configuration
Upgrade policy
Health and repai
Identity
Properties
Percentage CPU (Average)
0.22 %
Enable metric divide by instance count
Operator *
Greater than
Metric threshold to trigger scale action *
70
%
Duration (in minutes) *
10
Time grain (in mins)
1
Time grain statistic *
Average
Action
Operation *
Increase count by
Cool down (minutes) *
5
Instance count *
1
Add
toscale to
ces based on
e is above

Demo



Deploying to an Availability Zones

Deploying to an Availability Sets

Create Virtual Machine Scale Set

Summary

Availability zones distribute VMs across *Azure regions*

- SLA 99.99%

Fault domains are a rack of servers in an Azure datacenter

Update domains servers protect against normal maintenance

- Ie. Hardware updates

Availability sets group VMs to distribute across a single datacenter

- SLA 99.95%
- Must be assigned during deployment

Summary

Virtual machine scale sets are a group of load balanced VMs

- No additional costs
- Can be deployed across multiple update/fault domains
- Rules can be created based on performance metrics or schedules
- Understand how a rule may impact the number of instances scaled based on metrics