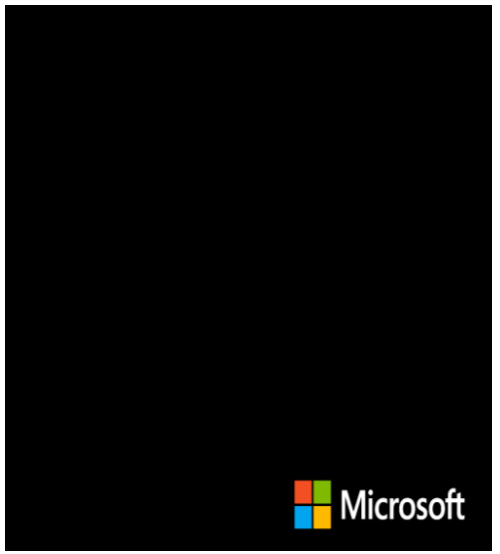
Compute Infrastructure

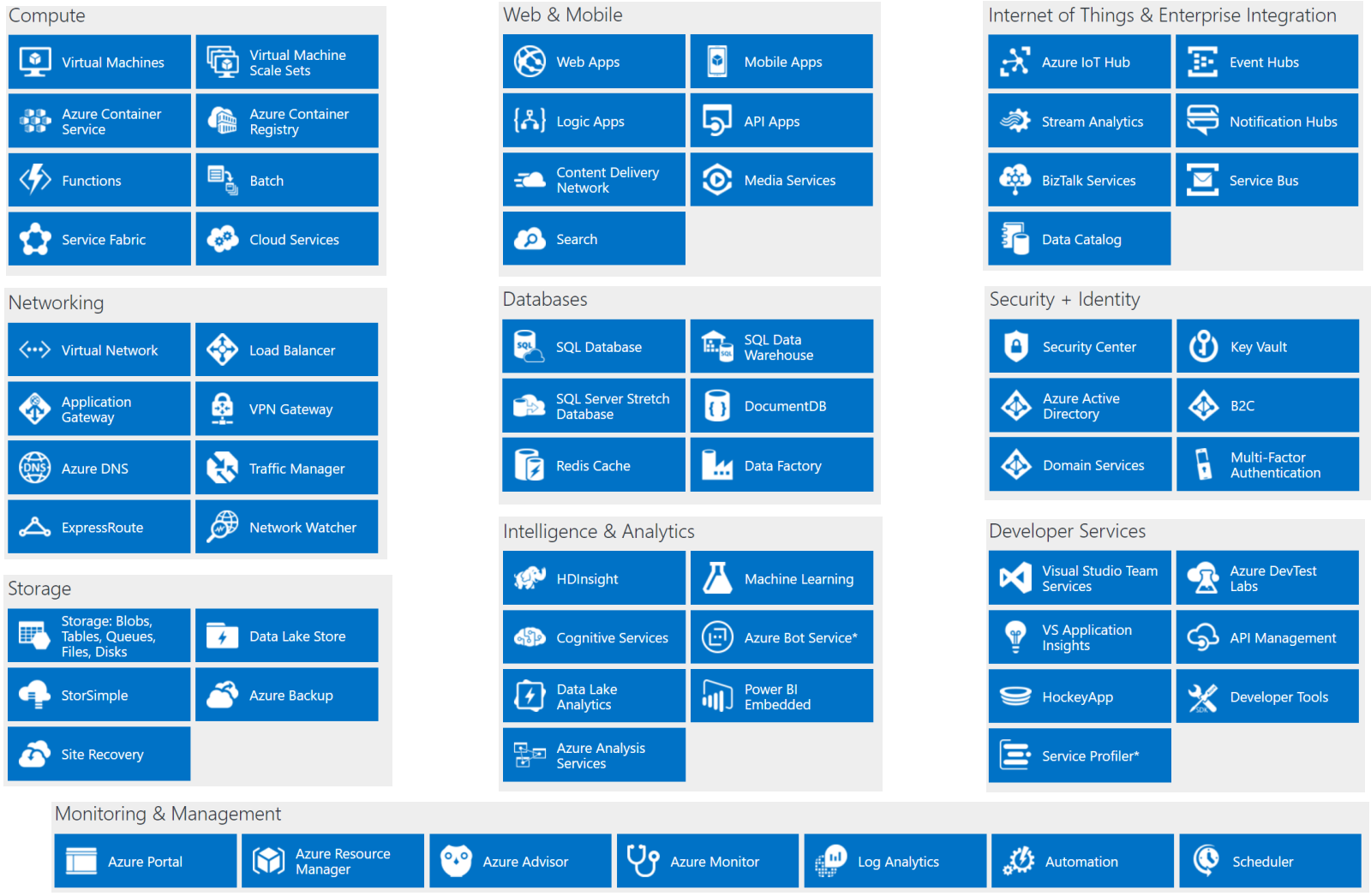


Dan Rey
Cloud Consultant
Technical Trainer | MCT

Design Compute Infrastructure (20-25%)

- Design solutions using virtual machines
  - Design VM deployments by leveraging availability sets, fault domains, and update domains in Azure; use web app for containers; design VM Scale Sets; design for compute-intensive tasks using Azure Batch; define a migration strategy from cloud services; recommend use of Azure Backup and Azure Site Recovery
- Design solutions for serverless computing
  - Use Azure Functions to implement event-driven actions; design for serverless computing using Azure Container Instances; design application solutions by using Azure Logic Apps, Azure Functions, or both; determine when to use API management service
- Design microservices-based solutions
  - Determine when a container-based solution is appropriate; determine when container-orchestration is appropriate; determine when Azure Service Fabric (ASF) is appropriate; determine when Azure Functions is appropriate; determine when to use API management service; determine when Web API is appropriate; determine which platform is appropriate for container orchestration; consider migrating existing assets versus cloud native deployment; design lifecycle management strategies
- Design web applications
  - Design Azure App Service Web Apps; design custom web API; secure Web API; design Web Apps for scalability and performance; design for high availability using Azure Web Apps in multiple regions; determine which App service plan to use; design Web Apps for business continuity; determine when to use Azure App Service Environment (ASE); design for API apps; determine when to use API management service; determine when to use Web Apps on Linux; determine when to use a CDN; determine when to use a cache, including Azure Redis cache
- Create compute-intensive application
  - Design high-performance computing (HPC) and other compute-intensive applications using Azure Services; determine when to use Azure Batch; design stateless components to accommodate scale; design lifecycle strategy for Azure Batch

Azure Platform

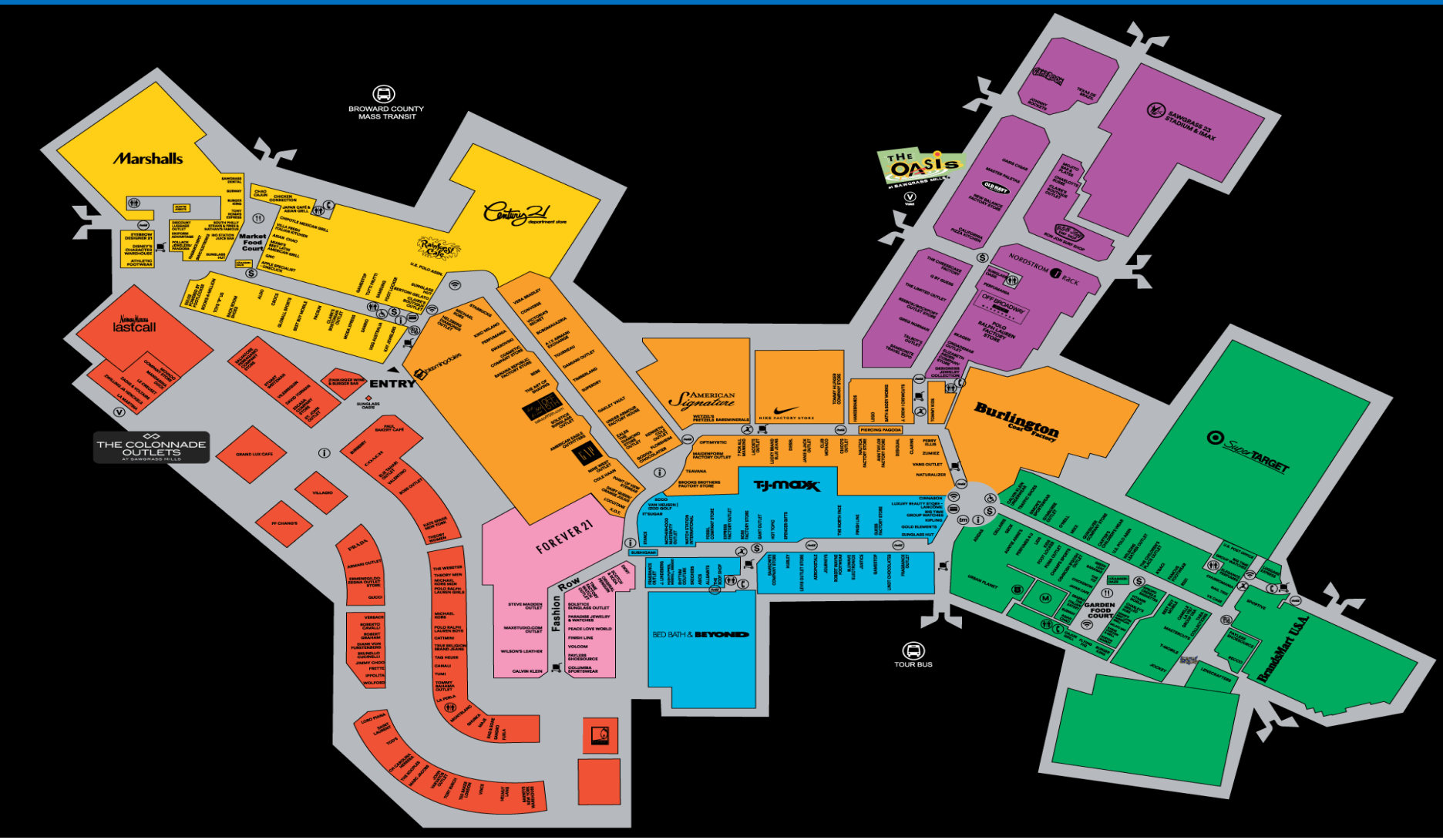


http://azureplatform.azurewebsites.net/

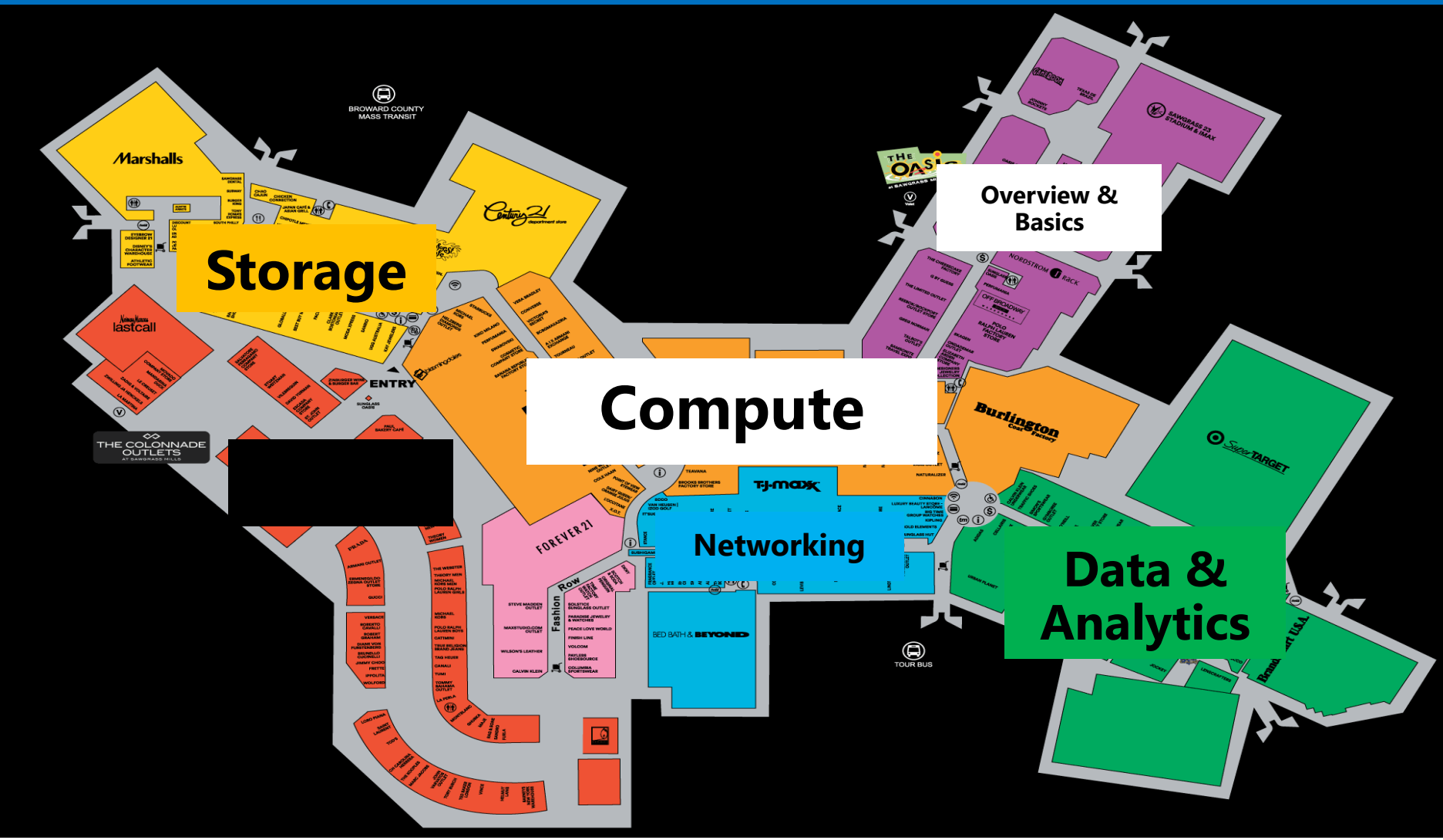
* Preview Services

#AzureArch @stilldrey

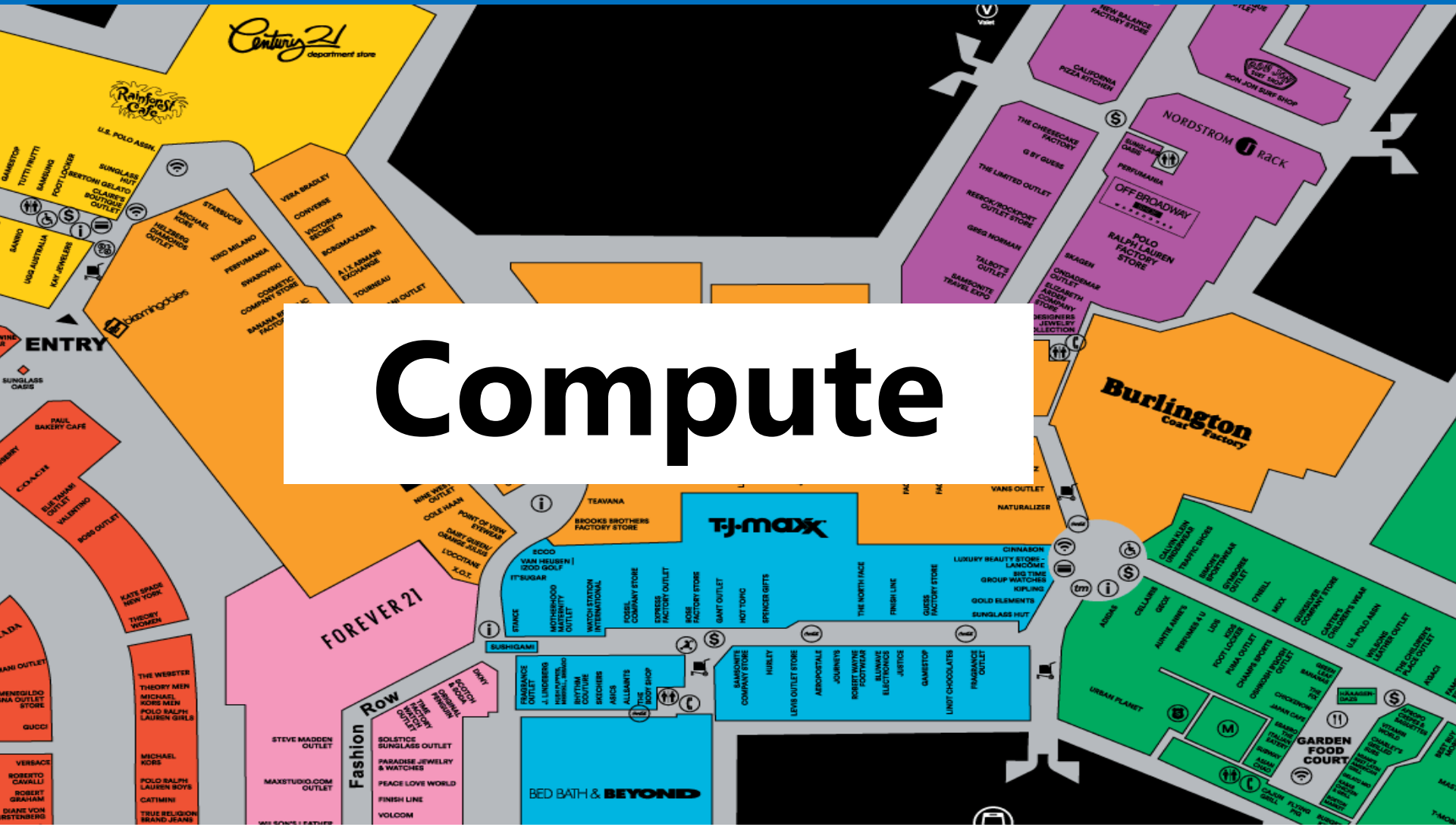
The Mall


BROWARD COUNTY MASS TRANSIT
Marshalls
Century 21 department store
Rainforest Cafe
Market Food Court
Last Call
ENTRY
THE COLONNADE OUTLETS
AT SAWGRASS MILLS
FOREVER 21
Fashion Row
American Signature
T.J.maxx
BED BATH & BEYOND
THE OASIS
AT SAWGRASS MILLS
SAWGRASS 23 STADIUM & MAX
NORDSTROM RACK
OFF BROADWAY SHOES
Burlington Coat Factory
Super TARGET
BrandsMart U.S.A.
GARDEN FOOD COURT
TOUR BUS

Azure


Overview & Basics
Storage
Compute
Networking
Data & Analytics
BROWARD COUNTY MASS TRANSIT
Marshalls
Century 21
lastcall
ENTRY
THE COLONNADE OUTLETS
FOREVER 21
T.J.maxx
Burlington
Super TARGET
BED BATH & BEYOND
TOUR BUS
Fashion Row

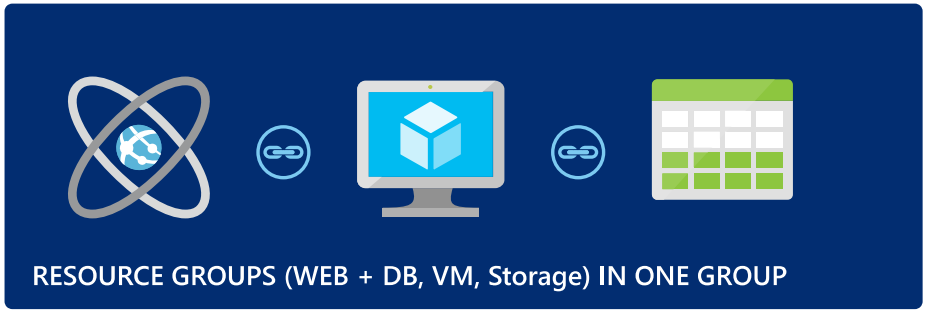
An area within the mall



Azure Resource Manager (ARM)

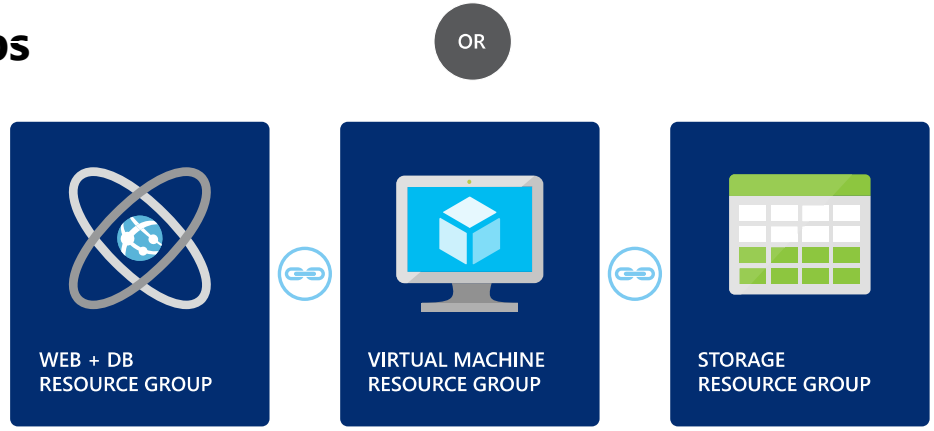
All services offered as **ARM resources**

- Consistent model for creating/managing Azure resources
- Based on declarative templates



Resources are managed in **resource groups**

- Deployed together
- Managed together
- Provides RBAC support



Azure Resource Manager Terminology

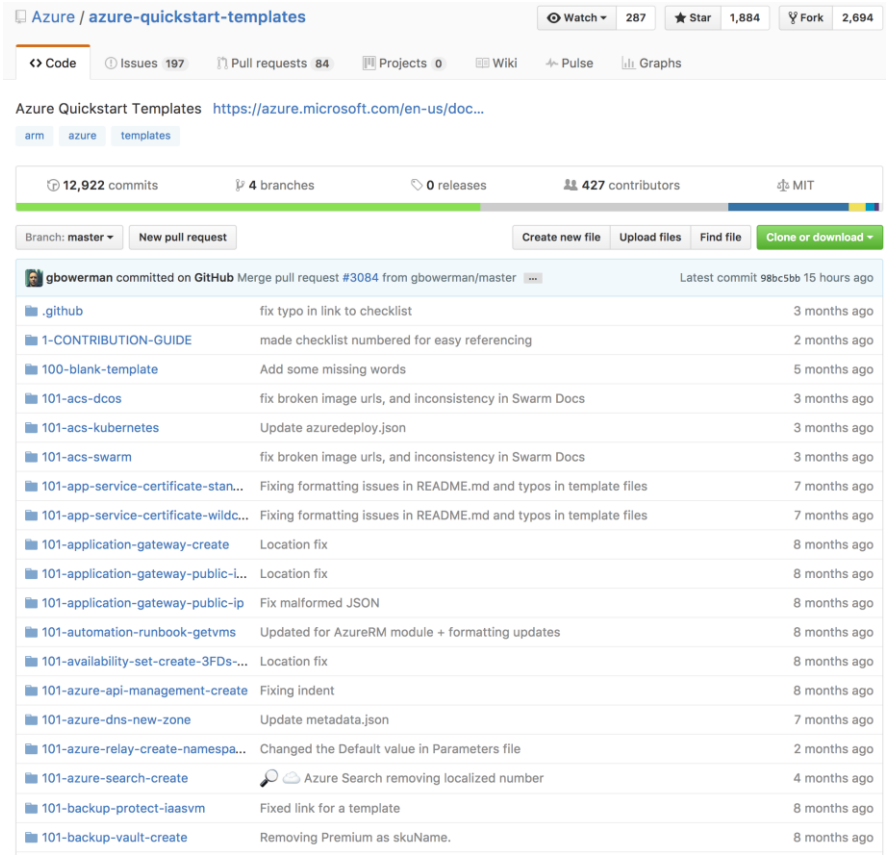
- Resource
  - A manageable unit of Azure infrastructure
  - Examples: virtual machine, load balancer, storage account
- Resource provider
  - A service that allows for creation and management of resources
  - Examples: Compute Resource Provider, Storage Resource Provider
- Resource manager template
  - A declarative template for describing a resource group

Reusable ARM templates

- A declarative template for describing a resource group
- Community offered or Build your own

- Learn about template here:

https://azure.microsoft.com/en-us/resources/templates/



Azure: The Power Of Choice

Compute

Virtual Machines

Container Service

Service Fabric

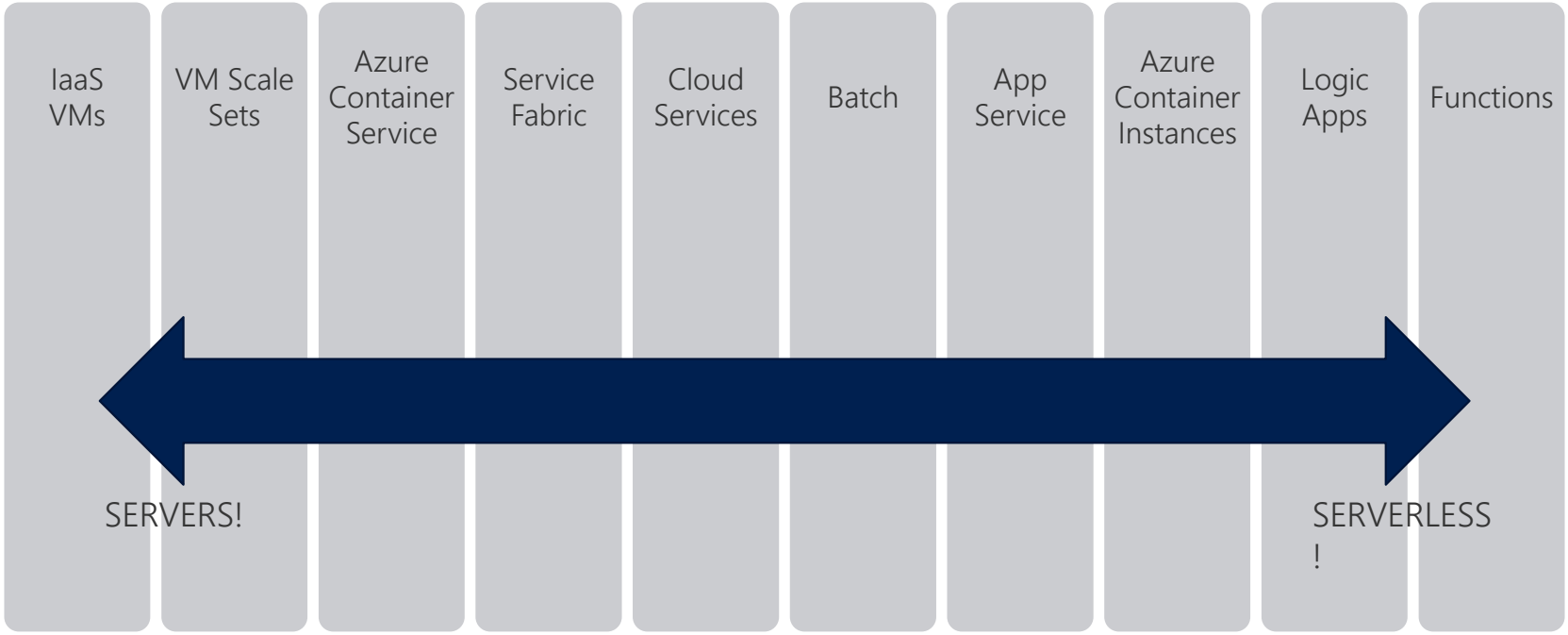
App Service

Functions

More Control ← → Focus on the App

Customer-managed (IaaS)	Platform-managed (PaaS)	Code-only (serverless)

Compute



https://aka.ms/comparecompute

Virtual Machines

Ubuntu, Red Hat, Windows, SUSE, CoreOS

DevOps Extensions with Chef and Puppet

Multiple sizes

Hundreds of items in marketplace

The Ultimate Driving Machine®

VEHICLES | BUILD YOUR OWN | SHOPPING TOOLS | Certified Pre-Owned | Insi

ALL SERIES | Sedans | Coupes | Convertibles | SAVs & Sports Wagons | Diesel | Hybrid | Electric

2 Series

$32,850 starting M SRP[1]

AVAILABLE IN
Coupe, Convertible

3 Series

$33,150 starting M SRP[1]

AVAILABLE IN
Sedan, Sports Wagon, Gran Turismo

4 Series

$41,650 starting M SRP[1]

AVAILABLE IN
Coupe, Gran Coupe, Convertible

5 Series

$50,200 starting M SRP[1]

AVAILABLE IN
Sedan, Gran Turismo

6 Series

$77,300 starting M SRP[1]

AVAILABLE IN
Coupe, Gran Coupe, Convertible, ALPINA Gran Coupe

7 Series

$81,300 starting M SRP[1]

AVAILABLE IN
Sedan

X Models

$34,800 starting M SRP[1]

AVAILABLE IN
Sports Activity Vehicle, Sports Activity Coupe

Z4

$49,700 starting M SRP[1]

AVAILABLE IN
Roadster

M Models

$63,500 starting M SRP[1]

AVAILABLE IN
Sedan, Coupe, Gran Coupe, Convertible, Sports Activity Vehicle, Sports Activity Coupe

BMW i

$42,400 starting M SRP[1]

Introducing the Latest Azure VM Line Up

Azure VM Sizes

Lowest Price

SSD Storage Fast CPUs

New generation of D family VMs

High memory and Large SSDs

New A-Series

Compute Intensive

NVIDIA GPUs K80 Compute

NVIDIA GPUs M60 Visualization

Fastest CPU IB Connectivity

Large SSDs

SAP Large Instances

Deep Learning NVIDIA P40s

New gen of NC NVIDIA P100s

New generation of D family

High memor

VM Sizes

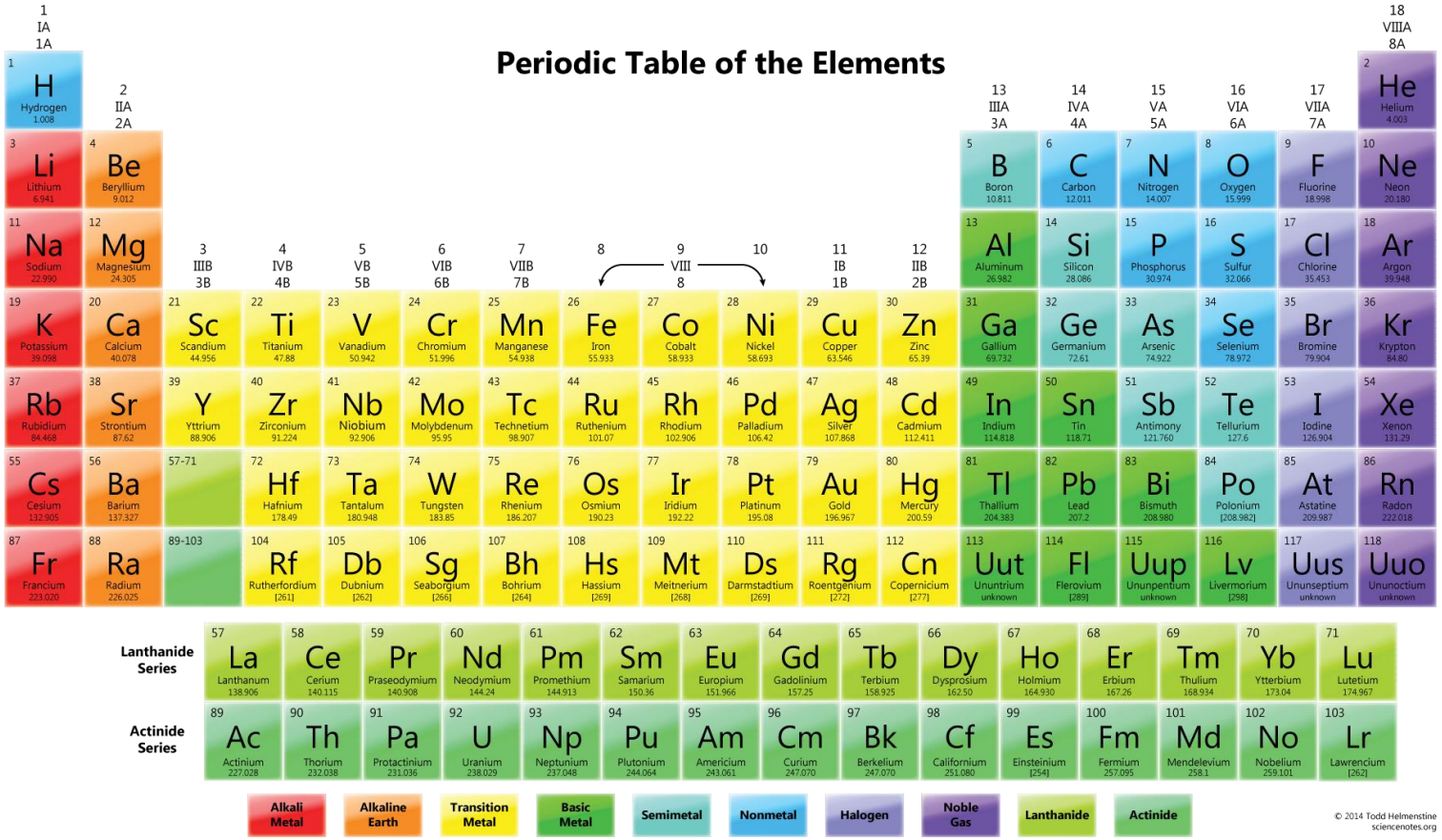
Type	Sizes	Description
General purpose	B (Preview), Dsv3, Dv3, DSv2, Dv2, DS, D, Av2, A0-7	Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers.
Compute optimized	Fs, F	High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers.
Memory optimized	Esv3, Ev3, M, GS, G, DSv2, DS, Dv2, D	High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.
Storage optimized	Ls	High disk throughput and IO. Ideal for Big Data, SQL, and NoSQL databases.
GPU	NV, NC	Specialized virtual machines targeted for heavy graphic rendering and video editing. Available with single or multiple GPUs.
High performance compute	H, A8-11	Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA).

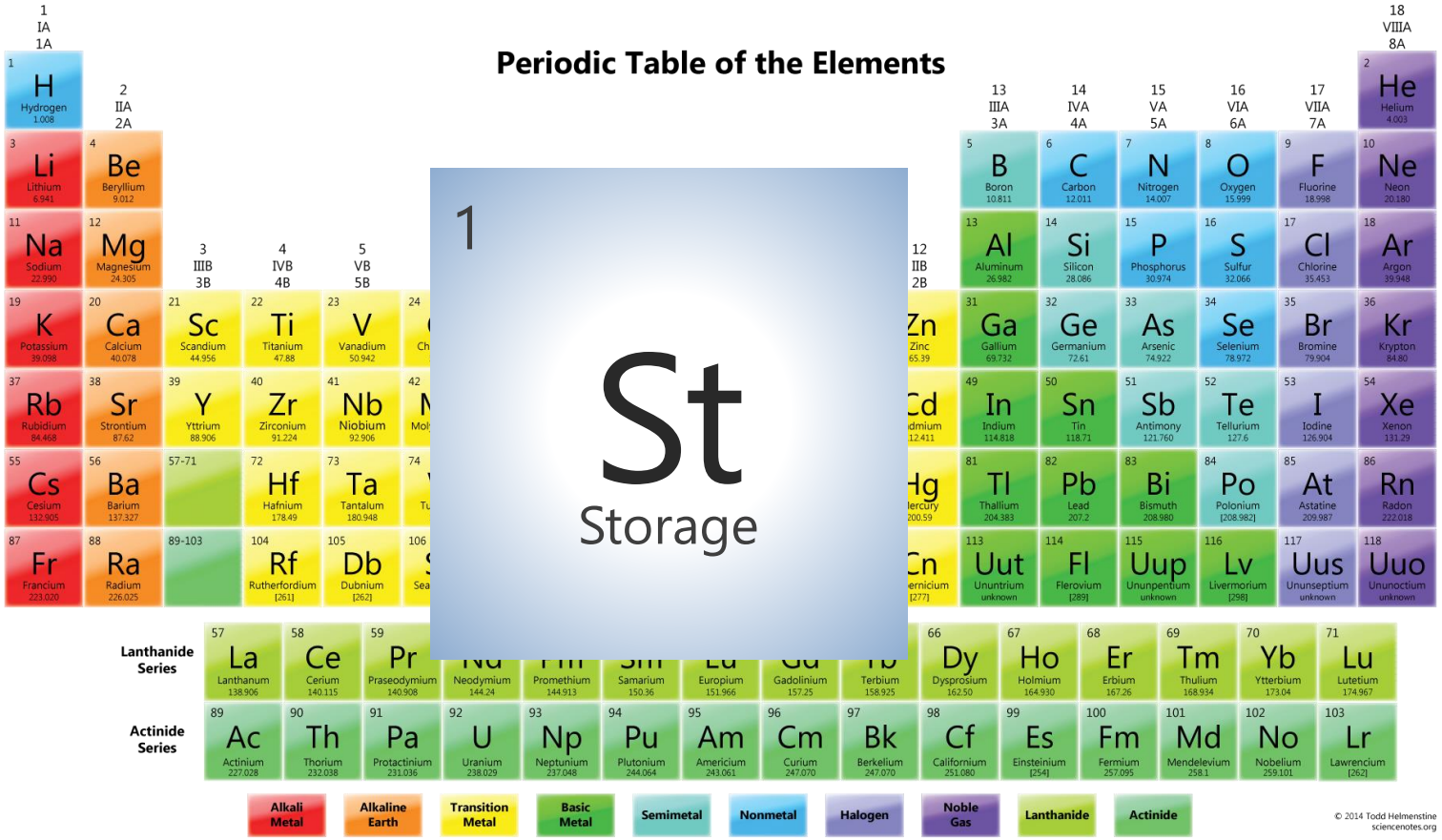
VM Azure Compute Unit (ACU)

- Provide a way of comparing compute (CPU) performance across Azure SKUs
- Not all Azure Cores are created equal
  - A1 Core != F1 Core
- Compare compute (CPU) performance across SKUs.

SKU Family	ACU \ vCPU
A0	50
A1-A4	100
A5-A7	100
A1_v2-A8_v2	100
A2m_v2-A8m_v2	100
A8-A11	225*
D1-D14	160
D1_v2-D15_v2	210 - 250*
DS1-DS14	160
DS1_v2-DS15_v2	210-250*
D_v3	160-190* **
Ds_v3	160-190* **
E_v3	160-190* **
Es_v3	160-190* **
F1-F16	210-250*
F1s-F16s	210-250*
G1-G5	180 - 240*
GS1-GS5	180 - 240*
H	290 - 300*
L4s-L32s	180 - 240*
M	160-180**

Periodic Table of the Elements




Periodic Table of the Elements
1
St
Storage
Lanthanide Series
Actinide Series
Alkali Metal
Alkaline Earth
Transition Metal
Basic Metal
Semimetal
Nonmetal
Halogen
Noble Gas
Lanthanide
Actinide
© 2014 Todd Helmenstine
sciencenotes.org

VM Disks

- OS Disk (attached via SATA)
  - VHD based
  - Persists
  - Separate storage cost
- Temporary Disk
  - Doesn't persist - SSD
  - No Separate storage cost
- Data Disk (SCSI)
  - VHD based
  - Persists
  - Separate storage cost
  - Current Max Data Disk Size: 4095GB

Premium vs Standard (SSD vs HDD) Comparision

	Azure Premium Disk	**Azure Standard Disk**
Disk Type	Solid State Drives (SSD)	Hard Disk Drives (HDD)
Overview	SSD-based high-performance, low-latency disk support for VMs running IO-intensive workloads	HDD-based cost effective disk support for Dev/Test VM scenarios
Scenario	Production and performance sensitive workloads	Dev/Test, non-critical, Infrequent access
Disk Size	P4: 32 GB (Managed Disks only) P6: 64 GB (Managed Disks only) P10: 128 GB P20: 512 GB P30: 1024 GB P40: 2048 GB P50: 4095 GB	Unmanaged Disks: 1 GB – 4 TB (4095 GB) Managed Disks: S4: 32 GB S6: 64 GB S10: 128 GB S20: 512 GB S30: 1024 GB S40: 2048 GB S50: 4095 GB
Max Throughput per Disk	250 MB/s	60 MB/s
Max IOPS per Disk	7500 IOPS	500 IOPS

VM Recommendations

Premium Storage for Production Workloads (Storage SLAs)

Choose a VM Size that works with premium storage for production

Use Managed Disks over Unmanaged Disks

Scaling Up/Down is just resizing the VM

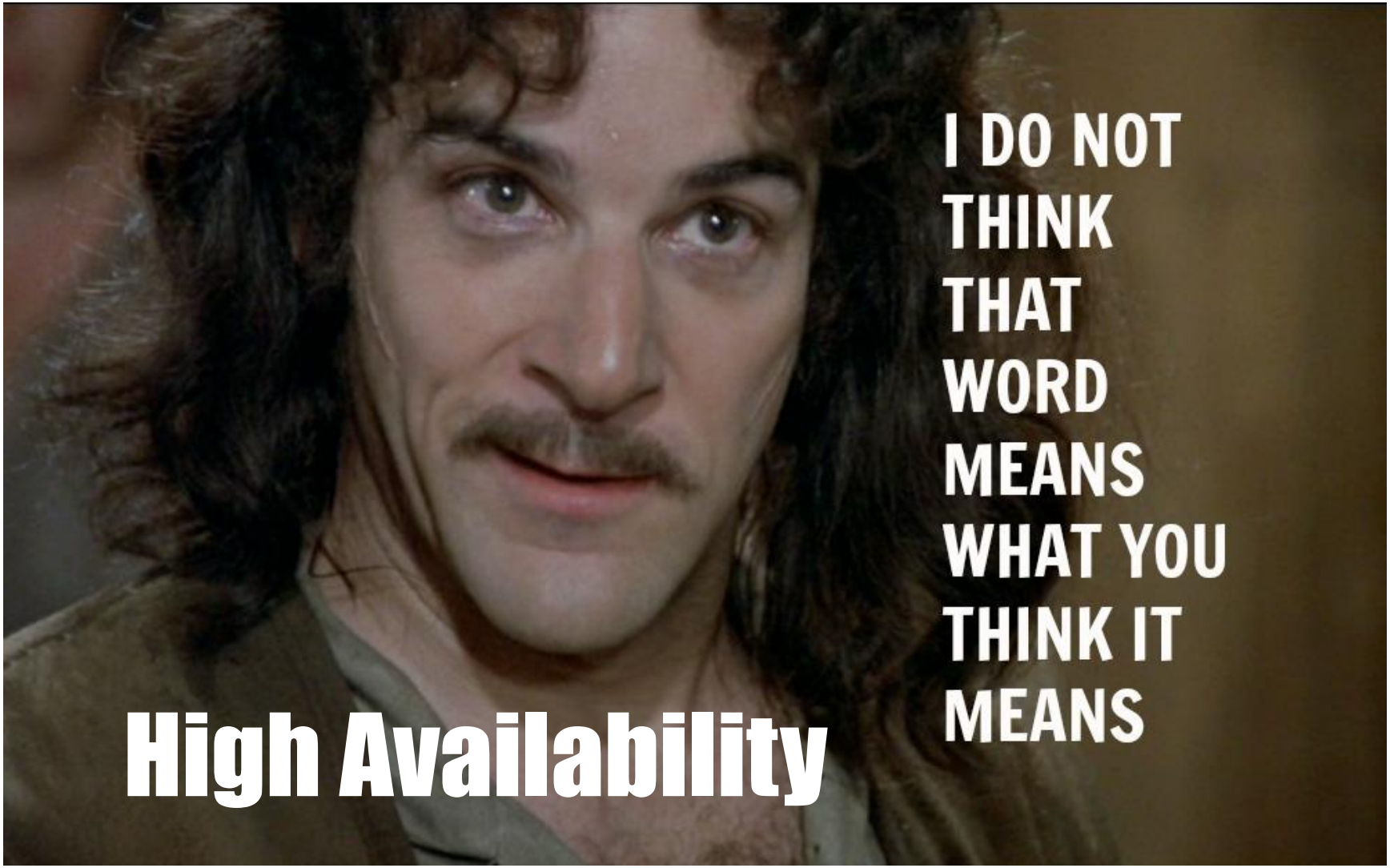
Scaling In/Out – the VMs should be in an availability set

Use VM reboot logs to determine if VM was rebooted by planned maintenance

Use snapshots to prevent accidental data loss

Enable VM diagnostics for production (includes boot diagnostics)

Stopped VMs are still charged for use. VMs need to be **deallocated** to stop charges. **Stopping through OS does not deallocate! Stop with portal or CLI.**


I DO NOT
THINK
THAT
WORD
MEANS
WHAT YOU
THINK IT
MEANS
High Availability

High Availability & Disaster Recovery
in Azure

- High Availability
  - Availability within a single Azure region or datacenter*
  - Expectation is little or no downtime (99.x % uptime)
- Disaster Recovery
  - Recover into a secondary datacenter if outage in primary datacenter
  - Acceptable downtime has a greater range
  - Quantified by Recovery Time Objective & Recovery Point Objective

Defining Disaster Recovery

- Disaster recovery is being able to recover your application into another Azure datacenter
- Keywords:
  - RTO: Recovery Time Objective
  - RPO: Recovery Point Objective
- Options range from:
  - Complete active/active deployment in multiple datacenters
  - Active/Passive deployment in multiple datacenters
  - Active deployment in one datacenter, with backups stored in 2nd datacenter

Understanding Azure VM Availability

Azure SLA guarantees no data loss, 99.9% uptime SLA*

- Subject to *un-planned maintenance* events due to physical failures
  - If VM becomes unavailable, Azure migrates VM and restarts in another host
  - ~10-15 minutes to complete this process

Understanding Azure VM Availability

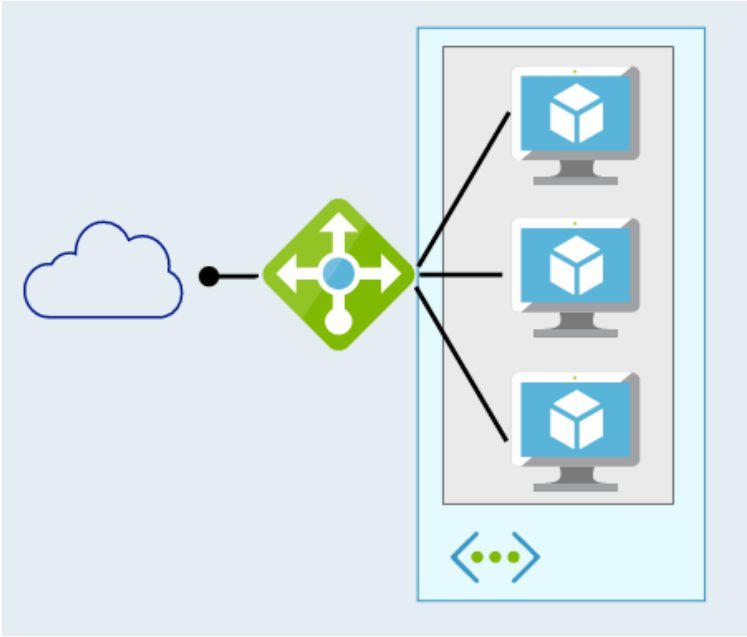
Azure SLA guarantees no data loss, 99.9% uptime SLA*

- Subject to *un-planned maintenance* events due to physical failures
  - If VM becomes unavailable, Azure migrates VM and restarts in another host
  - ~10-15 minutes to complete this process
- Subject to *planned maintenance* events due to host OS servicing
  - All VMs on host are shut down.
  - Host OS is serviced and rebooted
  - All VMs on host are restarted
  - ~10-15 minutes to complete this process
- Subject to in-memory *planned maintenance* events
  - All VMs on host are paused, Host patched, VMs un-paused. 30 seconds downtime

Defining High Availability

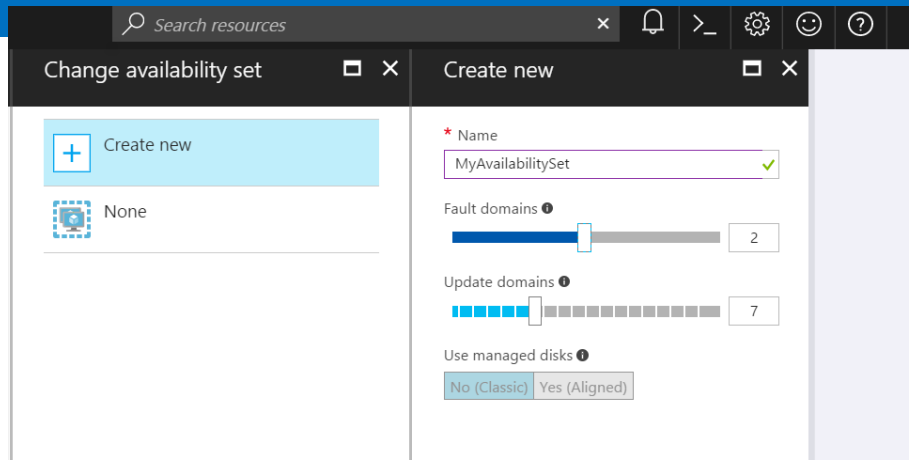
- Multiple VMs can be configured in an “availability set”
- Workload is load balanced across the VMs

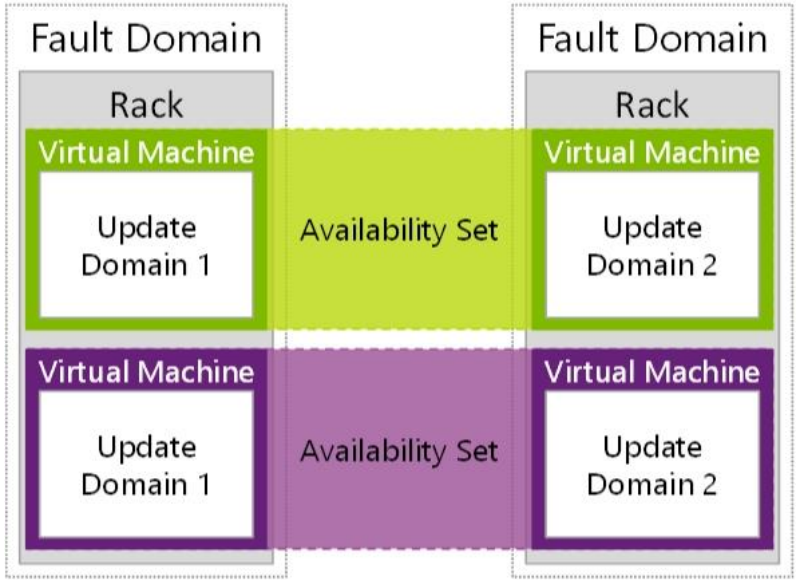
- Azure SLA: 2 (or more) VMs in Availability Set:
  - 99.95% (<22 min downtime p/month)
  - Includes
    - Planned downtime due to host OS servicing
    - Unplanned downtime due to physical failures
  - Doesn’t include servicing of guest OS or software inside (e.g. SQL)



Availability Sets

- Availability Sets are for Unplanned & Planned Maintenance
  - Fault Domains (2 default, some regions allow 3)
  - Upgrade Domains (5 default, 1-20 allowed)
- Front with Load Balancer, App Gateway





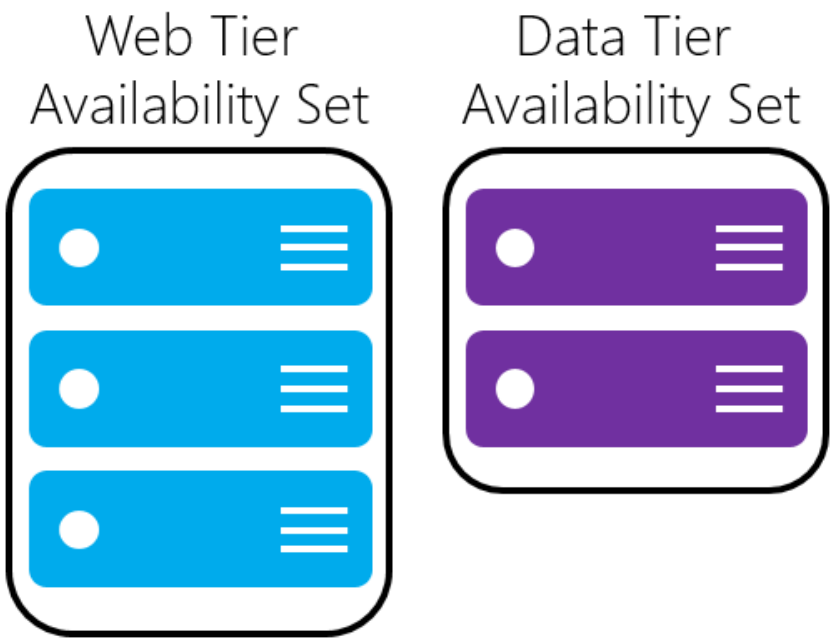
Availability Sets

- Do NOT put a single VM in an Availability Set

- Example - for an application,
  - Place front-end virtual machines in the same availability set

  AND

  - Data-tier virtual machines in their own availability set

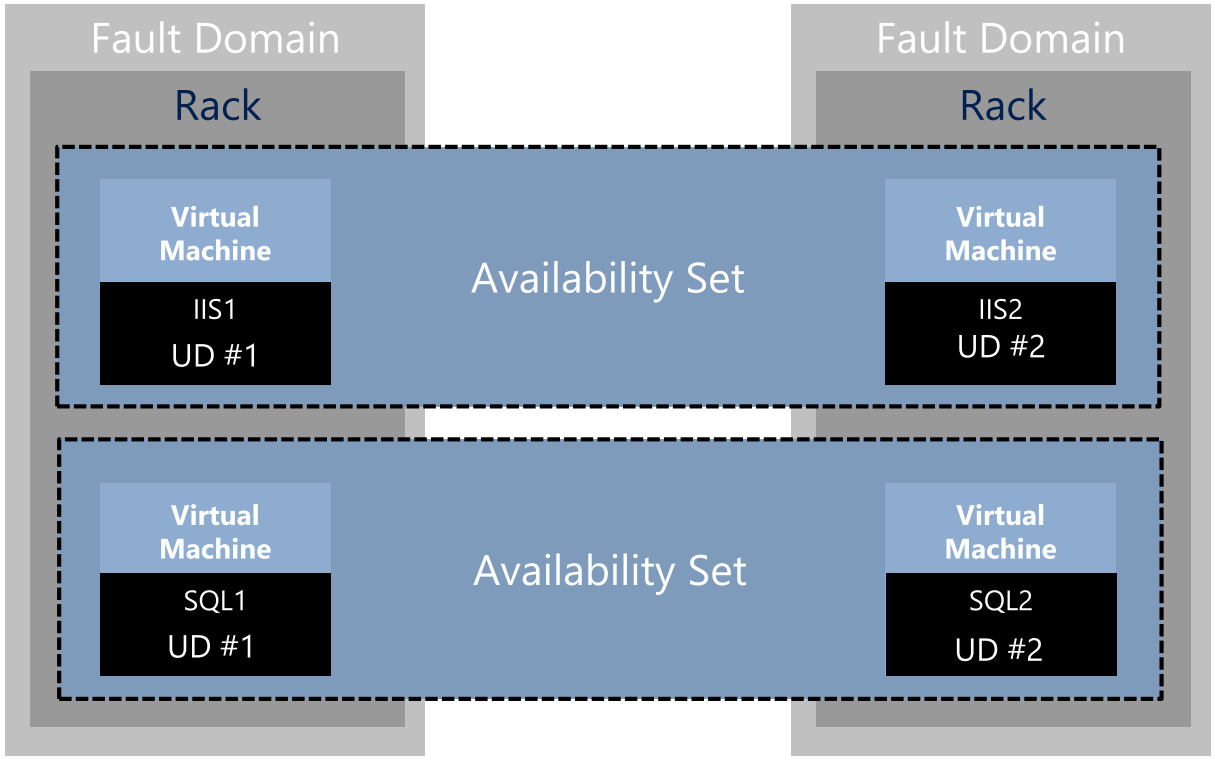


Availability Sets: How Do We Do It?
Fault and Update Domains

- Fault Domains
  - Represent groups of resources anticipated to fail together
  - i.e. Same rack, same server
  - Fabric spreads instances across fault at least 2 fault domains when two or more virtual machines are placed in an availability set
- Update Domains
  - Represents groups of resources that will be updated together
  - Host OS updates honor service update domains
  - Specified in service definition
  - Default of 5 (up to 20)
- Azure spreads role instances across Update Domains and Fault Domains

Virtual Machine Availability Sets

Update Domains are honored by host OS updates



VM Scale Sets

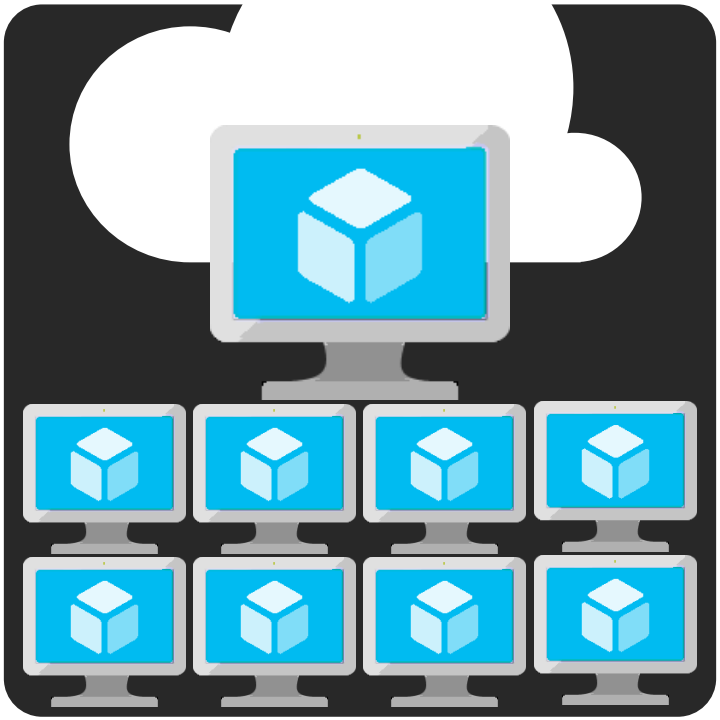
High performance provisioning of 1000+ VMs

Auto-configuration at scale

Auto-scale based on schedule and resource metrics

Easy updates at scale

Simple Portal Integration



Why VM Scale Sets?

- Manually **scale** with 'capacity' property
- **Autoscale** with host metrics (MDM pipeline) or diagnostic extensions
- Small buy-in: Deploy/manage sets of 0->100 identically configured VMs
- **Guest OS patching**: Patching primitives allow manually triggered rolling upgrades
- **High-availability** – implicit availability set with 5 FDs/5 UDs

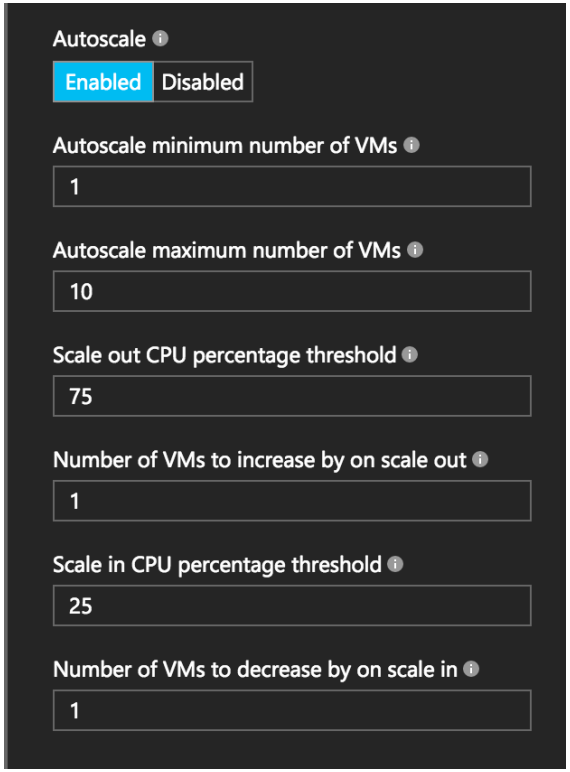
```
244     {
245       "type": "Microsoft.Compute/virtualMachineScaleSets",
246       "name": "[parameters('vmssName')]",
247       "location": "[parameters('resourceLocation')]",
248       "apiVersion": "[variables('computeApiVersion')]",
249       "dependsOn": [
250         "storageLoop",
251         "[concat('Microsoft.Network/loadBalancers/', variables('loadBalancerName'))]",
252         "[concat('Microsoft.Network/virtualNetworks/', variables('virtualNetworkName'))]"
253       ],
254       "sku": {
255         "name": "[parameters('vmSku')]",
256         "tier": "Standard",
257         "capacity": "[parameters('instanceCount')]"
258       },
259       "properties": {
260         "overprovision": "true",
261         "upgradePolicy": {
262           "mode": "Manual"
263         },
264         "virtualMachineProfile": {
```

Availability Sets vs Scale Sets

- Avail Set: Multiple different VMs (image, size, etc); managed separately
- Scale Set: Large count of the same VMs; managed together
- Scale set: Reliable rapid provisioning and scale utilizing similarity of the VMs

Autoscale with VM Scale Sets

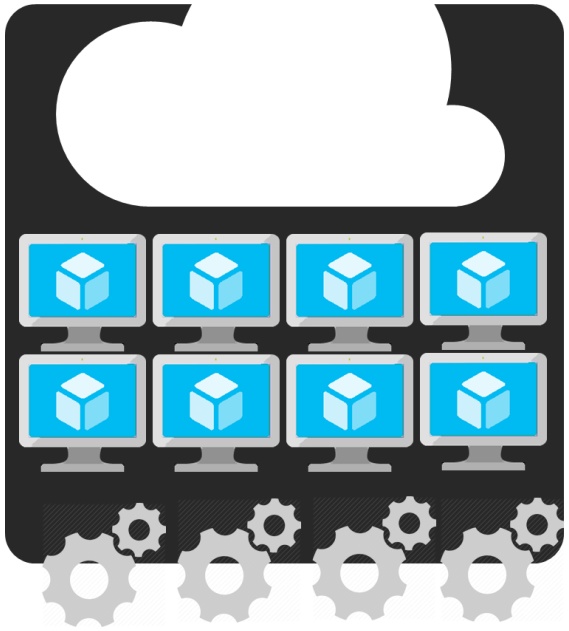
- Define Max – Min VMs
- Define trigger and action rules
- Standard audit / email notifications
- Define webhooks for custom notifications and actions (e.g. runbooks)



VM scale set app deployment models

Marketplace	Off the shelf solutions.
VM Extensions	Full control over app lifecycle management.
Custom data/unattend	Install custom app independently of external network.
Configuration manager	Centrally managed app installation, credentials & maintenance.
Containerized	Abstract app management from infrastructure. Cloud/DC agnostic.
Custom image	Small self-contained apps. Fast deploy. Immutable build, test, deploy pipelines.

Azure Batch



Compute pools for job processing

Automatic scaling and regional coverage

Linux and Windows

Automatically recover failed tasks

Input/Output handling

Low-Priority (discounted) option

Design microservices-based solutions

Azure: The Power Of Choice

Application Hosting (today)

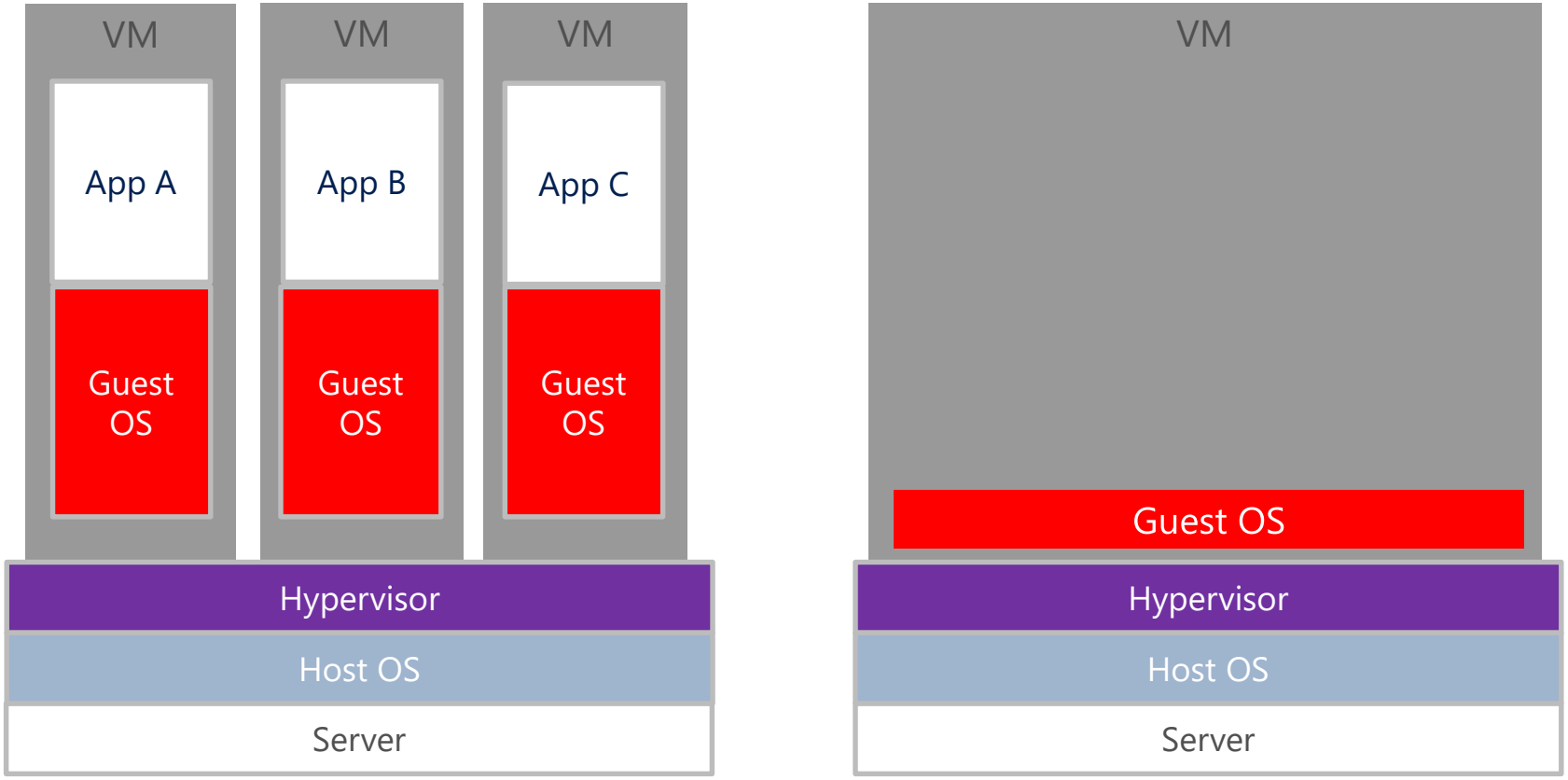
Virtual Machines

Containers

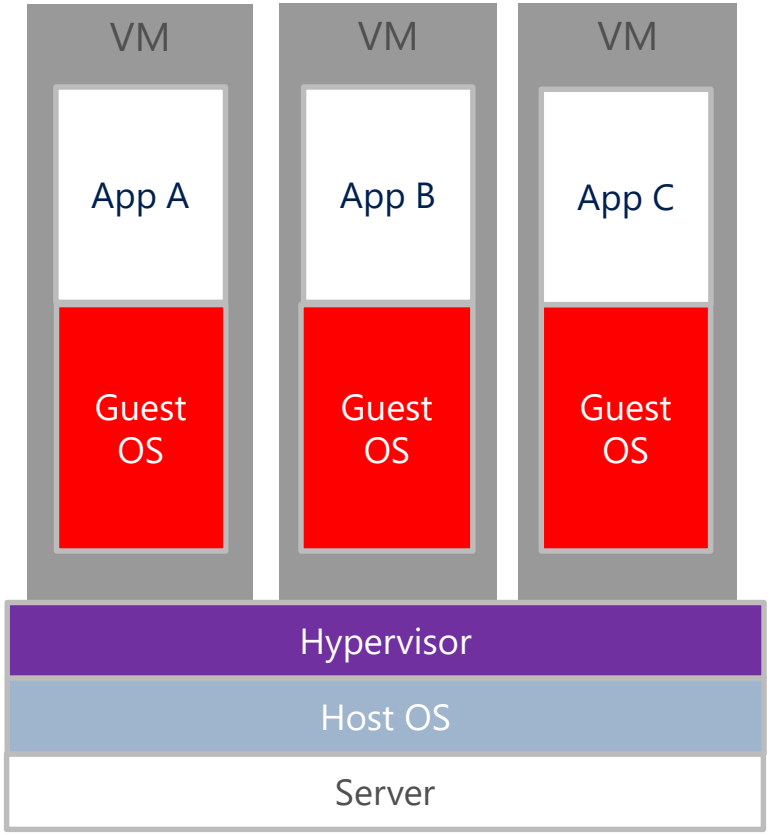
Customer-managed
(IaaS)

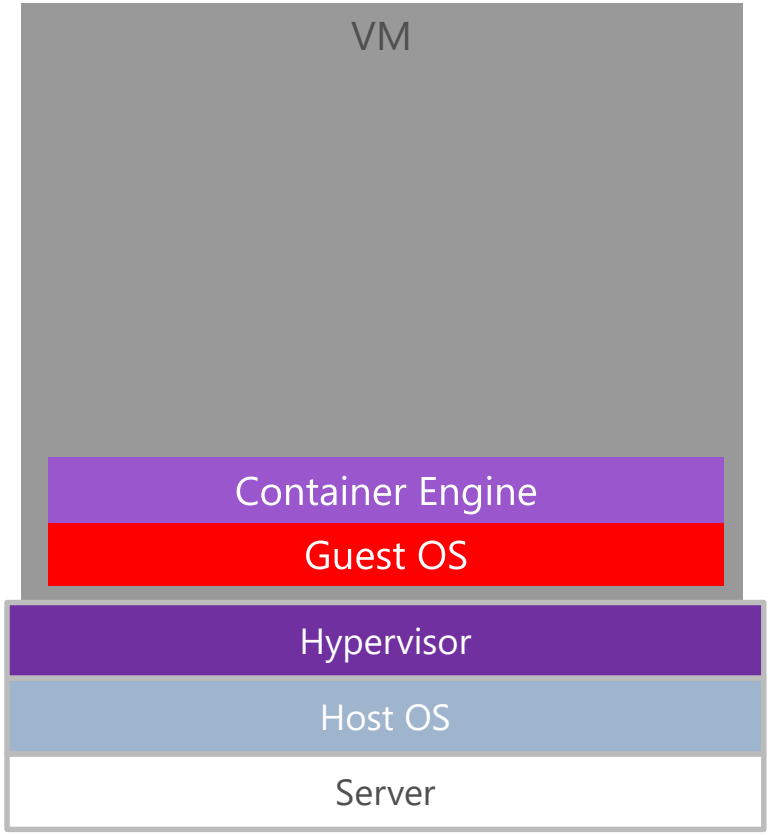
- What is a Container?

What is a Container?

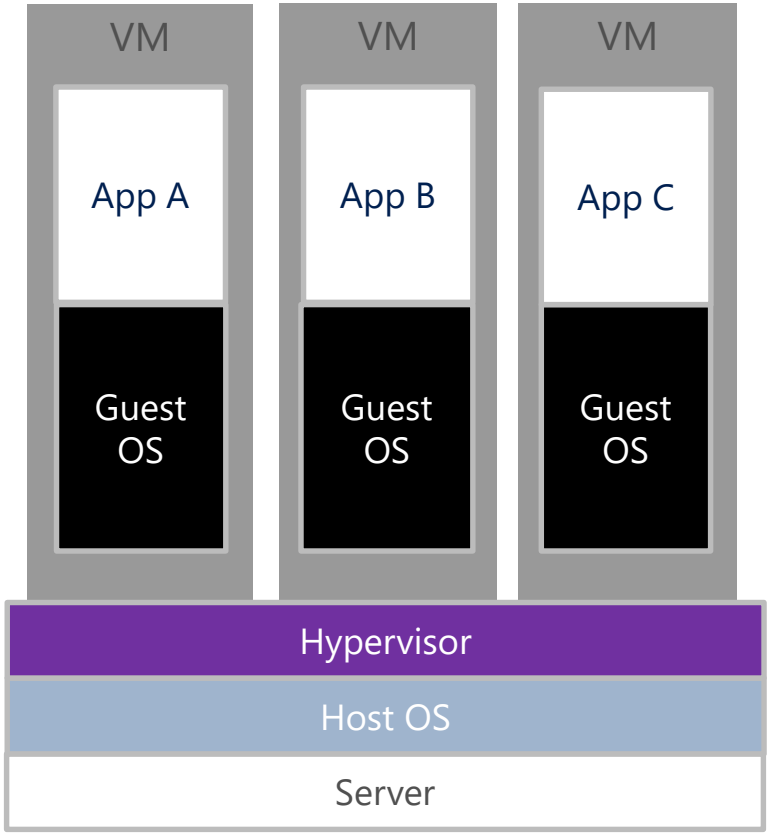


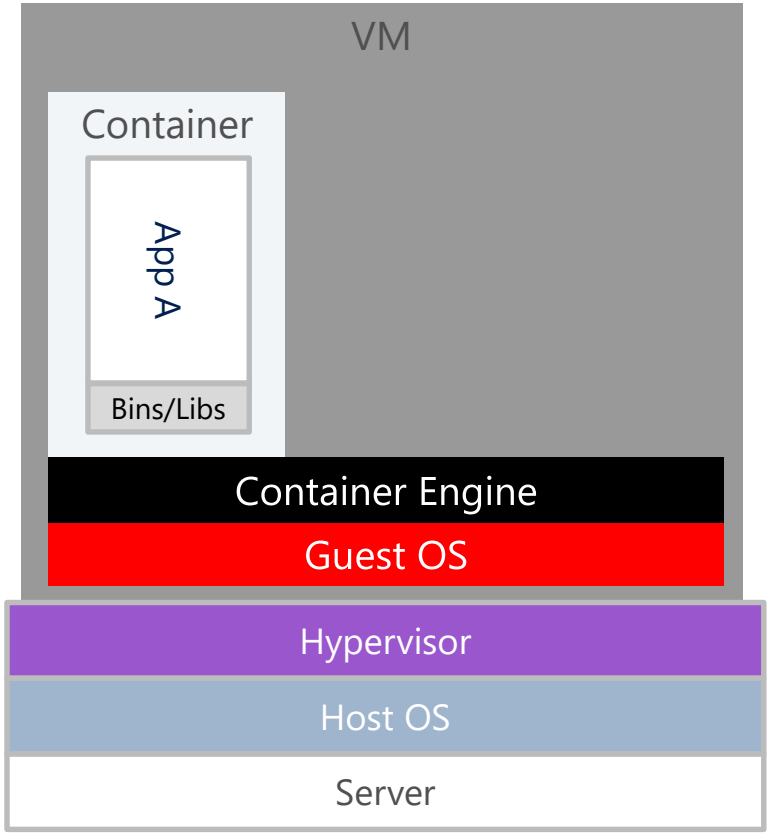
What is a Container?



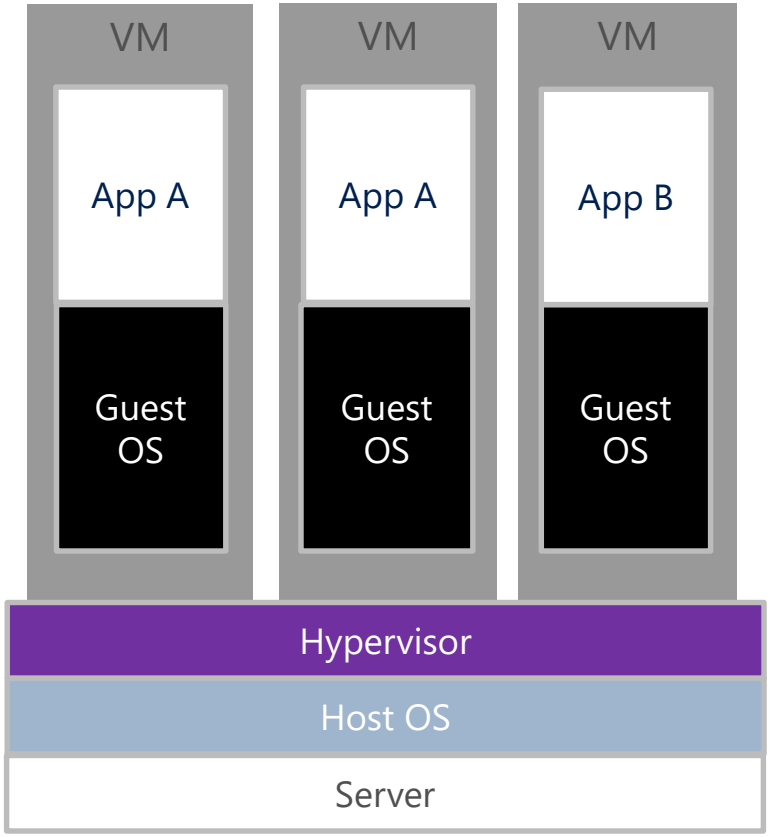


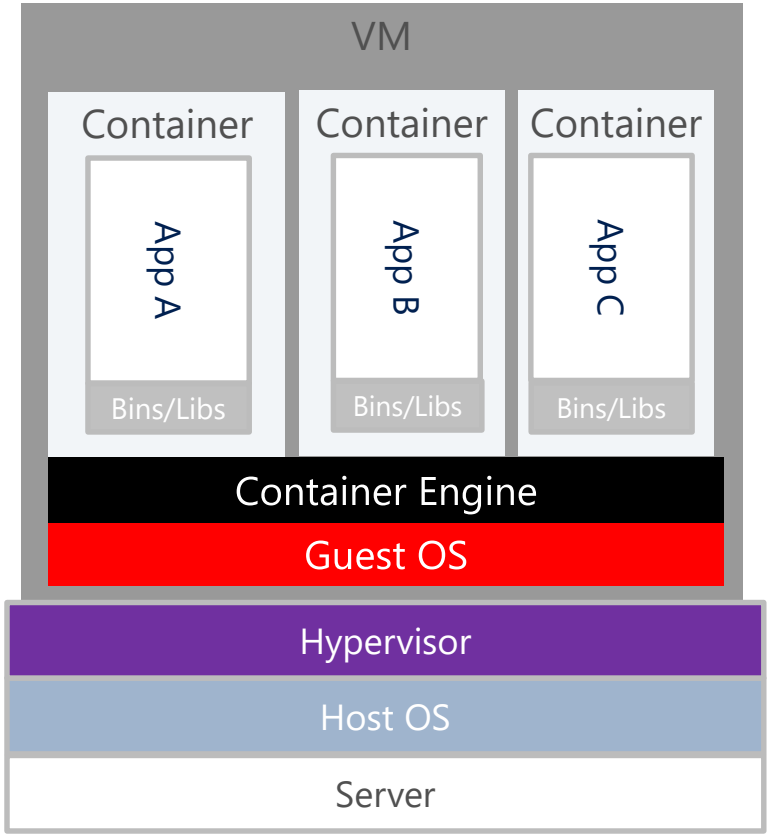
What is a Container?





What is a Container?





- Why?

- Lightweight

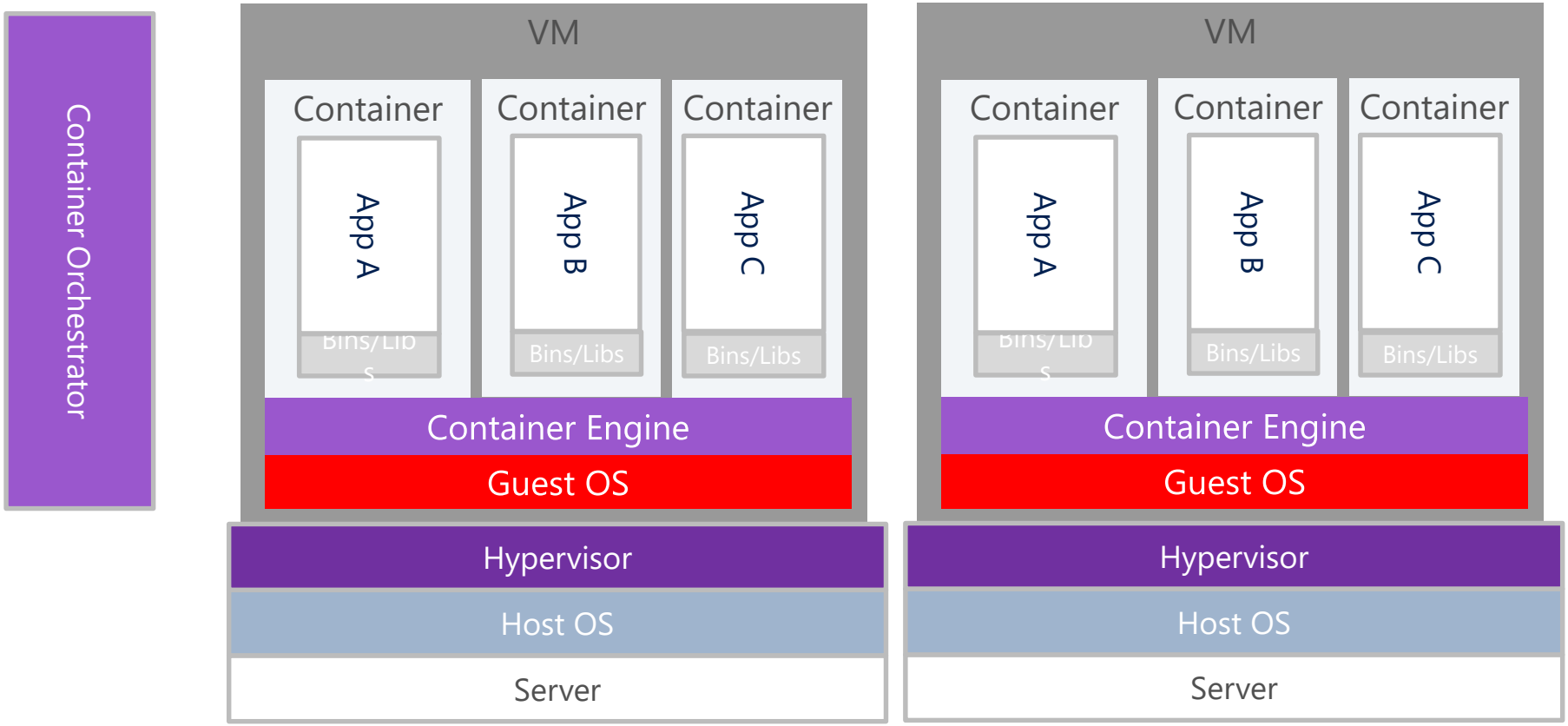
- Portable


works on
my
machine

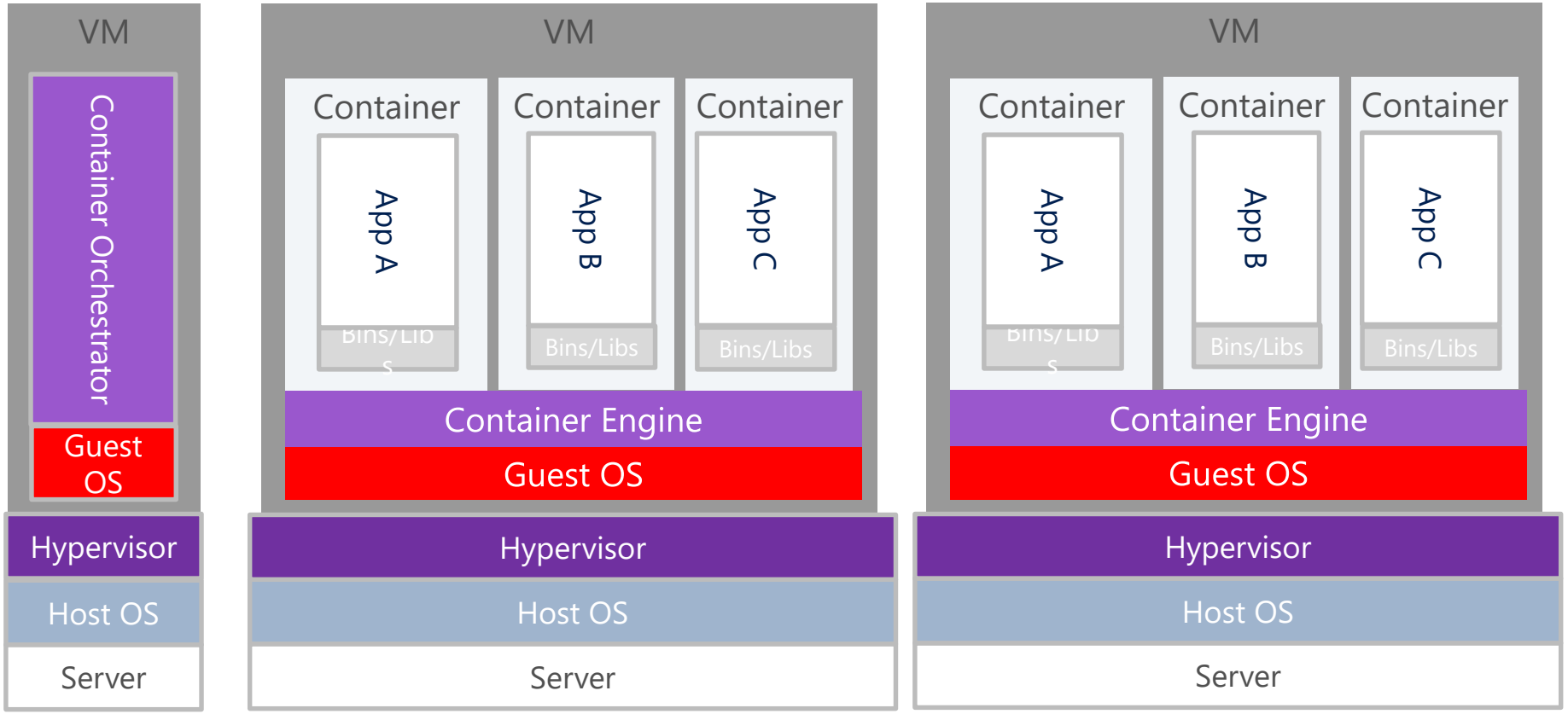
Ok, I've got containers. Now what?

What is a Container Orchestrator?

What is a Container Orchestrator?



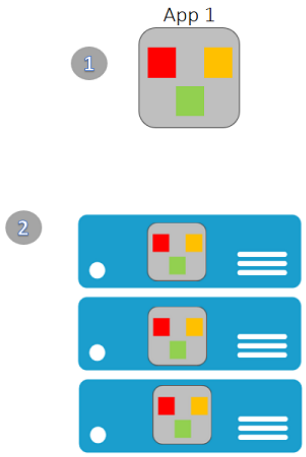
What is a Container Orchestrator?



What is the point of all of this?!

Microservices 101

Monolithic application approach



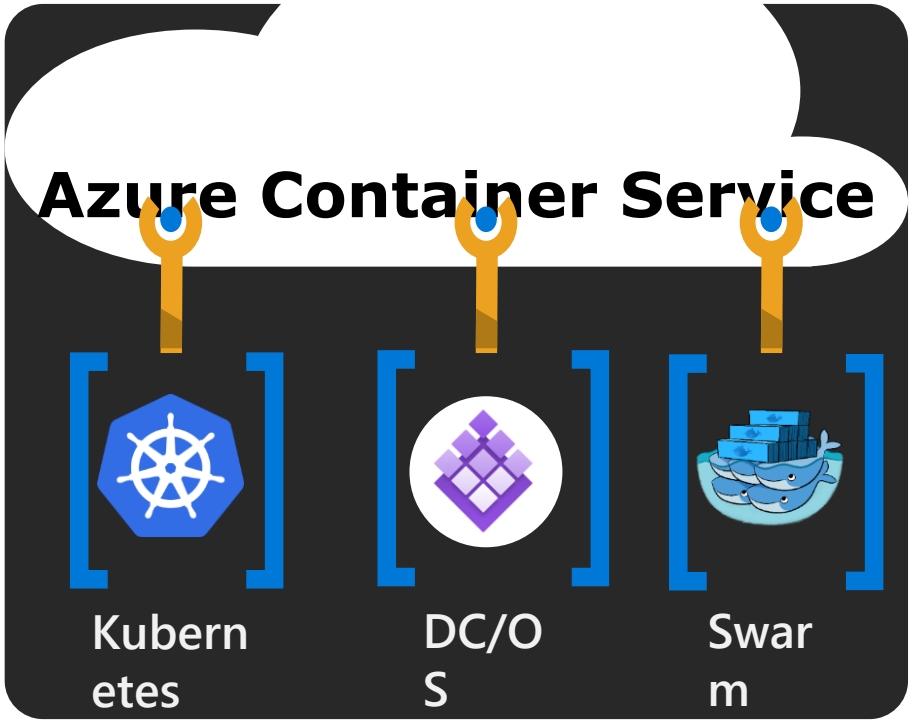
Azure Container Service

Making containers in Azure EZ!

- Standard Docker tooling & API Support
- Streamlined provisioning of K8S, DCOS, and Docker Swarm
- Linux & Windows Server Containers
- Azure & Azure Stack

ACS --> AKS (Preview)

- Focus on Kubernetes
- Managed Orchestrator



Azure Container Instances (Preview)

- Simplest and easiest way to run individual containers in the cloud
- No VM management
- Per-second billing with customized resource requests
- Linux and Windows Server containers

Azure: The Power Of Choice

Application Hosting (today)

Virtual Machines

Containers

Service Fabric

Customer-managed
(IaaS)

Platform-managed
(mIaaS/PaaS)

Microsoft Azure Service Fabric

A platform for reliable, hyperscale, microservice-based applications