



AWS Academy

North America Talent Development Team

Available AWS Certifications

Professional

Two years of comprehensive experience designing, operating, and troubleshooting solutions using the AWS Cloud



Associate

One year of experience solving problems and implementing solutions using the AWS Cloud



Architect

Operations

Developer

Foundational

Six months of fundamental AWS Cloud and industry knowledge

Cloud Practitioner



Specialty

Technical AWS Cloud experience in the Specialty domain as specified in the exam guide



Solutions Architect Associate Certification will test your ability to...

- Effectively demonstrate knowledge of how to architect and deploy secure and robust applications on AWS technologies
- Define a solution using architectural design principles based on customer requirements
- Provide implementation guidance based on best practices to the organization throughout the life cycle of the project

Recommended AWS/IT Knowledge

- Hands-on experience on compute, networking, storage, and database AWS services
- Hands-on experience with AWS deployment and management services
- Ability to identify which AWS services meet a given technical requirement
- Knowledge of recommended best practices for building secure and reliable applications on the AWS platform
- An understanding of the basic architectural principles of building on AWS
- An understanding of the AWS global infrastructure
- An understanding of network technologies as they relate to AWS
- An understanding of security features and tools that AWS provides and how they relate to traditional services

About the exam

- Length - 130 minutes
- Questions – 65 (2 minutes/question)
- Multiple choice – select the best 1 answer
- Multiple response – select 2 or more correct answers
- Scoring – 720/1000 to pass

Domains

Domain	% of exam
1. Design Resilient Architectures	30%
2. Design High-Performing Architectures	28%
3. Design Secure Applications and Architectures	24%
4. Design Cost-Optimized Architectures	18%
	100%

Domains

- Domain 1: Design Resilient Architectures
- 1.1 Design a multi-tier architecture solution
- 1.2 Design highly available and/or fault-tolerant architectures
- 1.3 Design decoupling mechanisms using AWS services
- 1.4 Choose appropriate resilient storage

Domains

- Domain 2: Design High-Performing Architectures
- 2.1 Identify elastic and scalable compute solutions for a workload
- 2.2 Select high-performing and scalable storage solutions for a workload
- 2.3 Select high-performing networking solutions for a workload
- 2.4 Choose high-performing database solutions for a workload

Domains

- Domain 3: Design Secure Applications and Architectures
- 3.1 Design secure access to AWS resources
- 3.2 Design secure application tiers
- 3.3 Select appropriate data security options

Domains

- Domain 4: Design Cost-Optimized Architectures
 - 4.1 Identify cost-effective storage solutions
 - 4.2 Identify cost-effective compute and database services
 - 4.3 Design cost-optimized network architectures

Objective

- 👉 Understand the basics of Cloud computing
- 👉 Understand Amazon Web Services (AWS) and Core Services

Introduction

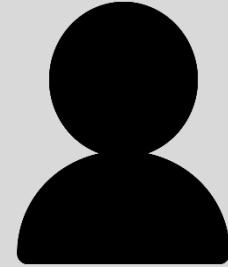
What is Cloud Computing?

As per NIST (National Institute of Standards and Technology) definition: Cloud computing is a model for enabling ubiquitous, convenient, on –demand network Access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

				
On-demand self-service	Broad network access	Resource pooling	Rapid elasticity	Measured service
No Human intervention to get additional resource	Access from Anywhere	Provider shares resources to customers	Get more resources quickly as needed	Pay only for what you consume

Can you think of examples of “clouds” that you use in your day-to-day?

Individuals



Cloud for personal use.
Mainly online services such
as webmail, on-demand-
content and cloud storage
solutions.



EVERNOTE

NETFLIX



Office 365



Businesses



Businesses use cloud for a range of tailored IT solutions
both as a part of their core business and for IT as a non-
core business process.



Google
Cloud Platform



amazon
web services



Microsoft

service**now**

vmware®

salesforce



Virtualization

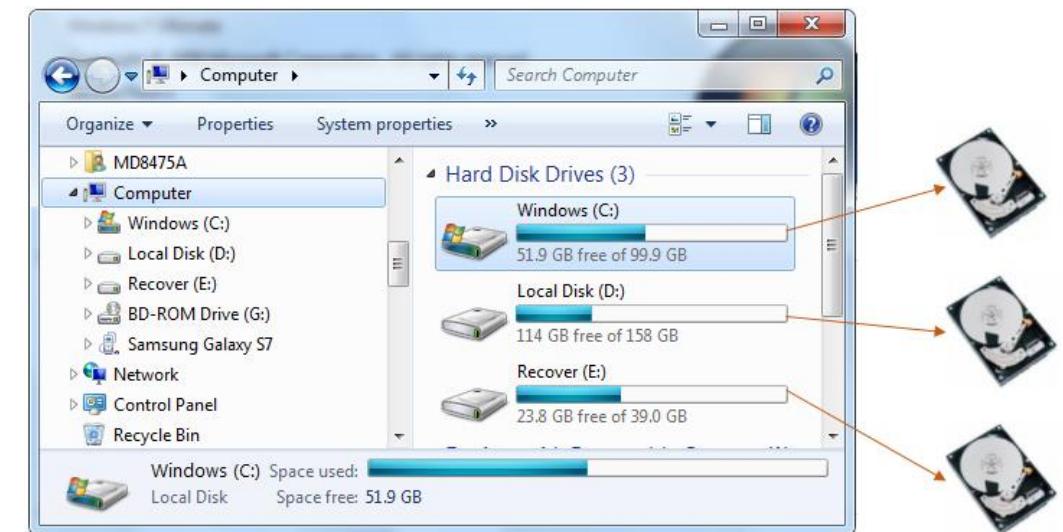
What is Virtualization?

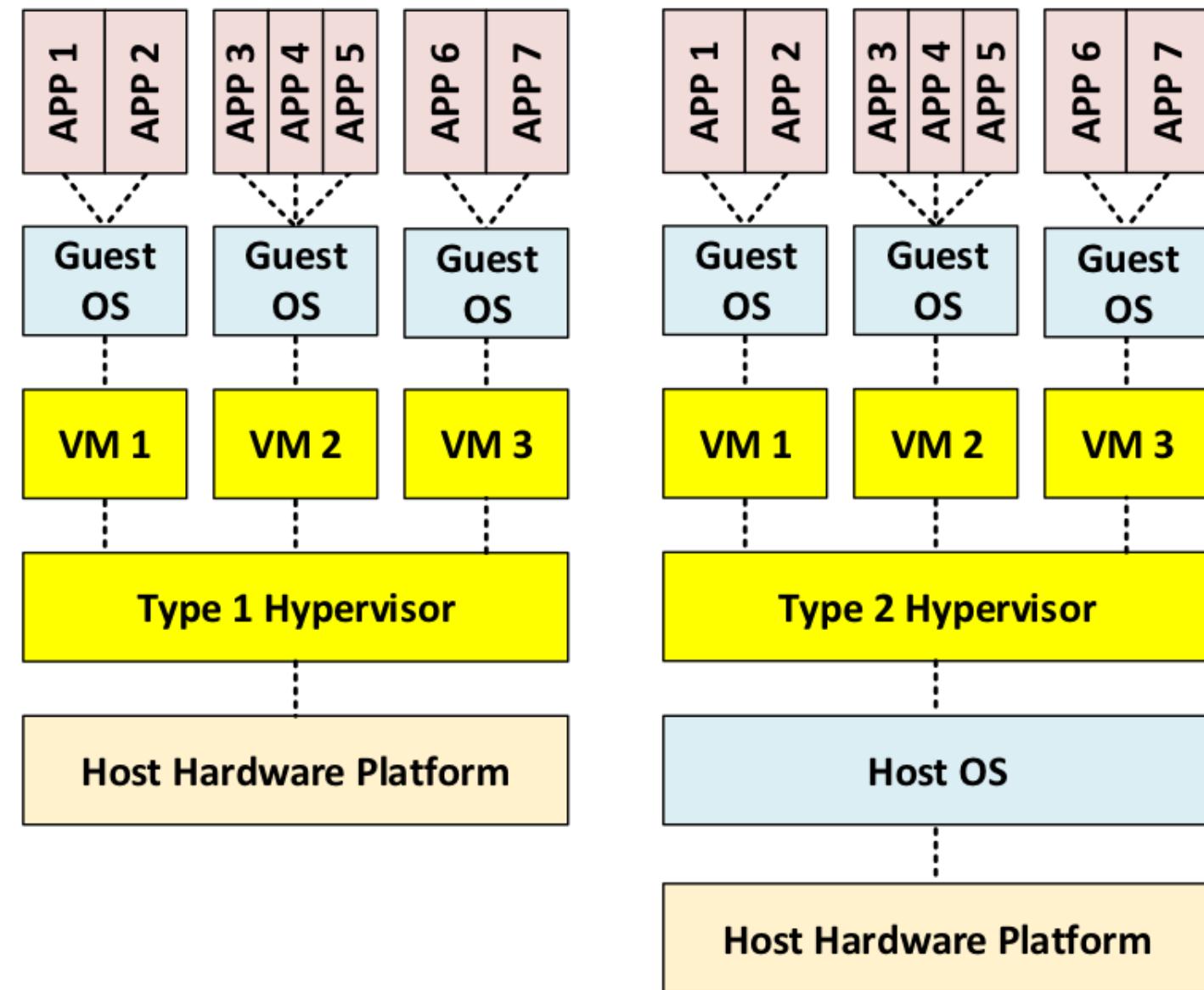
Virtualization refers to the **act of creating a virtual** (rather than actual) **version** of something, including **virtual computer hardware platforms, storage devices, and computer network resources**.

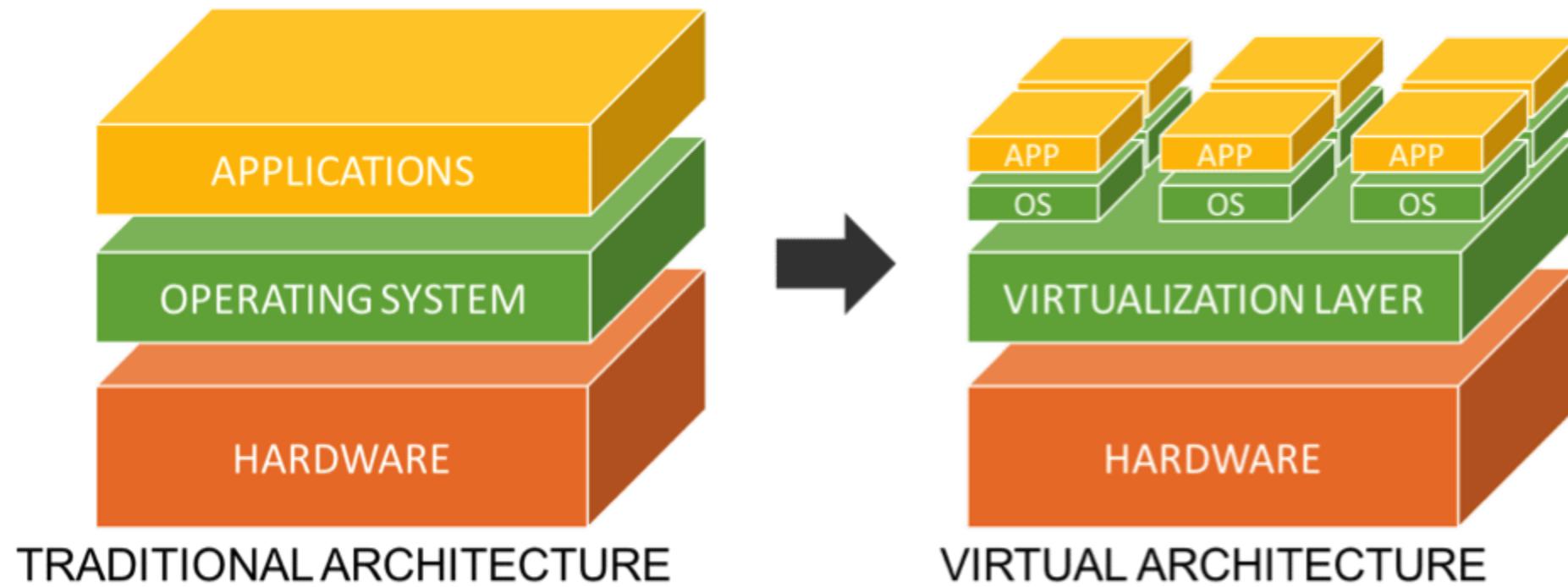


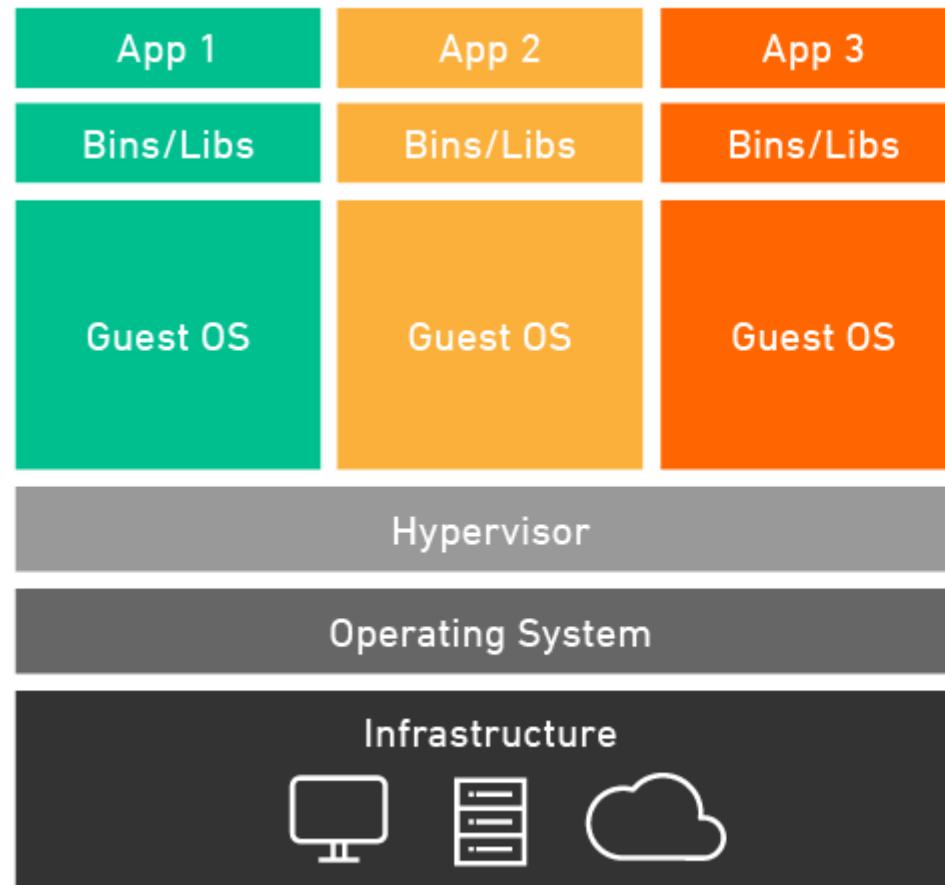
Best Example:

Separation of your hard drive into different drives (like C Drive, D Drive). While you may have only one hard drive, your system sees it as two, three or more different and separate segments

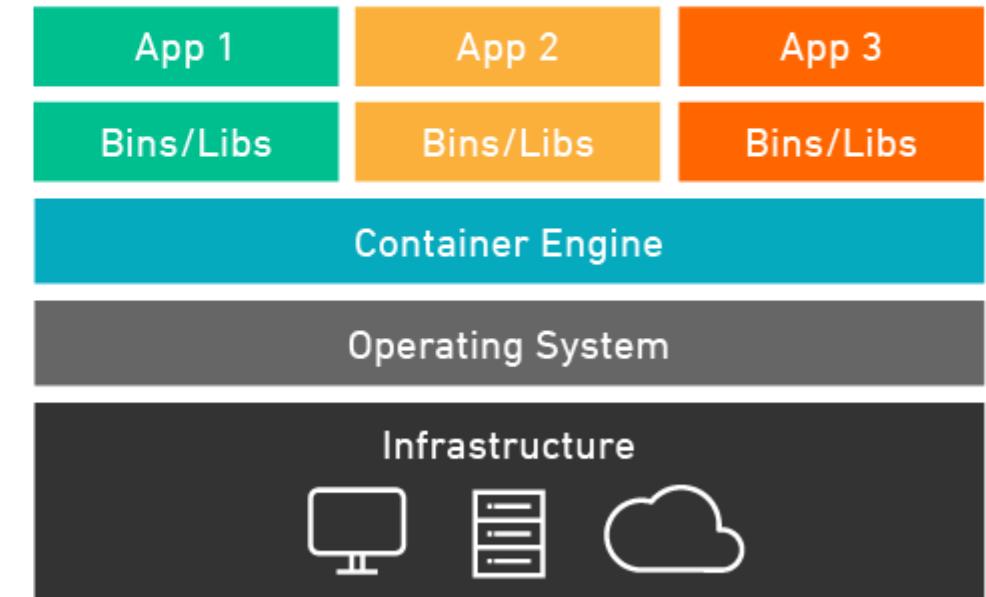






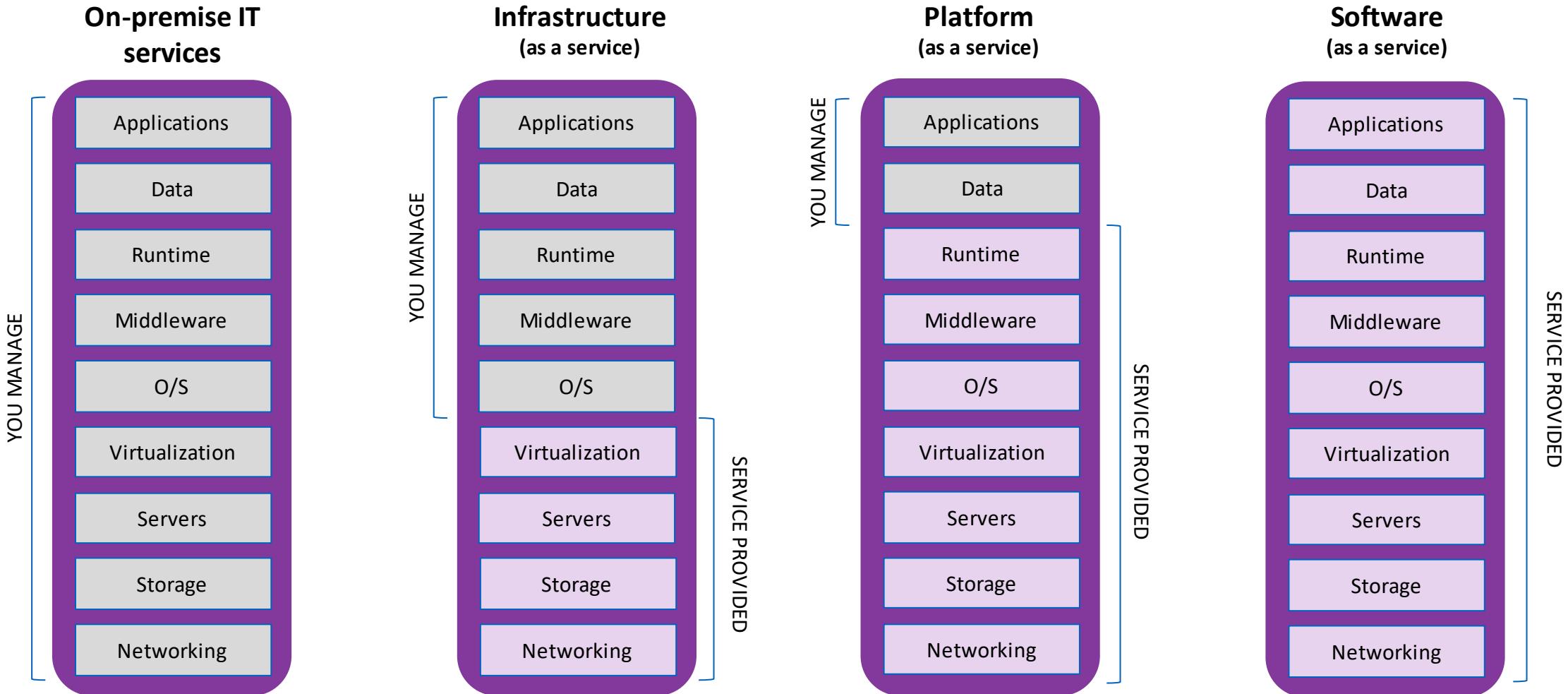


HYPERVERISOR-BASED VIRTUALIZATION

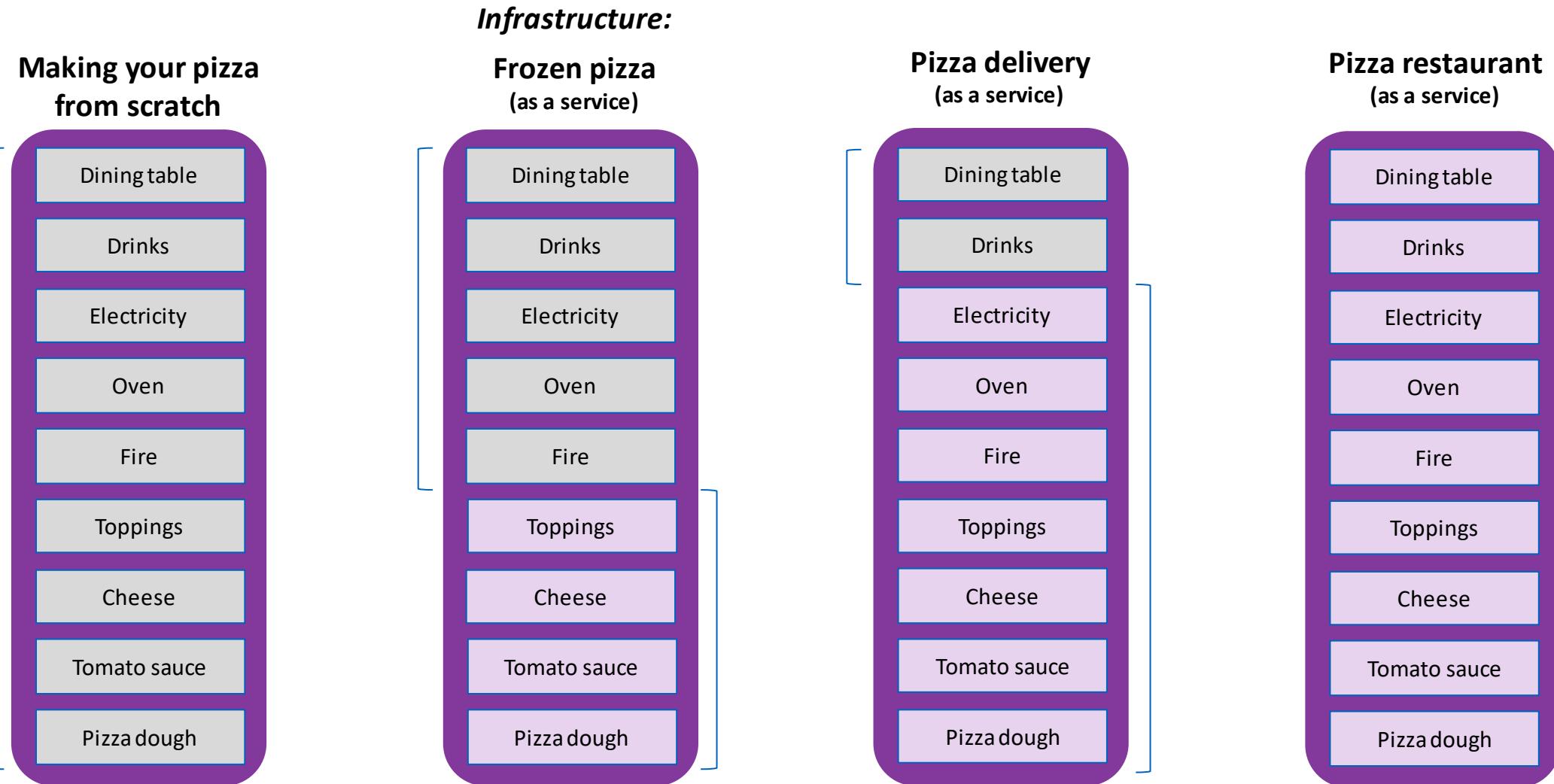


CONTAINER VIRTUALIZATION

What are the most common types of cloud services compared to standard, on-premise IT services?



Pizza as a Service



Cloud

Introduction to AWS



Objective

- 👉 Understand what is AWS Cloud and its history
- 👉 Understand basic of Global Infrastructure – Region, Availability Zones, Edge Locations
- 👉 AWS Services Overview
- 👉 AWS Shared security model

What is AWS?

- AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers millions of businesses in over 190 countries around the world.
- Benefits
 - More services and features than any other cloud provider
 - Services for storage, databases, networking, security, analytics, machine learning, IoT, and much more
 - Comprehensive portfolio of solutions that help you solve common problems and build faster using the AWS platform.
 - Every AWS Solution comes with detailed architecture, a deployment guide, and instructions for both automated and manual deployment.



<http://aws.amazon.com/what-is-cloud-computing/>
<http://aws.amazon.com/about-aws/>

What sets AWS apart?

	Security	Fine-grained control
	Service Breadth & Depth; pace of innovation	175+ services to support any cloud workload; rapid customer driven releases
	Experience: 1M+ customers	Building and managing cloud since 2006
	Global Footprint	767 Availability Zones within 24 geographic Regions, 2 Local Zone, 216 Points of Presence (205 Edge Locations and 11 Regional Edge Caches) in 84 cities across 42 countries.
	Machine Learning	More machine learning happens on AWS than anywhere else. Machine learning in the hands of every developer and data scientist.
	Ecosystem	Tens of thousands of APN partners. The AWS Marketplace offers 39 categories, and more than 7,000 software listings
	Enterprise leader	AWS positioned as a Leader in the Gartner Magic Quadrant for Cloud Infrastructure as a Service, Worldwide



Experience with Operational Reliability

- AWS have spent over a decade building the world's most reliable, secure, scalable, and cost-effective infrastructure.
- Service SLAs between 99.9% and 100% availability. Amazon S3 is designed for 99.999999999% durability.
- Availability Zones exist on isolated fault lines, flood plains, and electrical grids to substantially reduce the chance of simultaneous failure.
- The AWS Service Health Dashboard provides 24/7 visibility in the real-time operational status of all services around the globe.

Advantages of AWS Cloud



Six advantages of Cloud Computing



1. Trade capital expense for variable expense



2. Benefit from massive economies of scale



3. Stop guessing capacity



4. Increase speed and agility



5. Stop spending money running and maintaining data centers



6. Go global in minutes

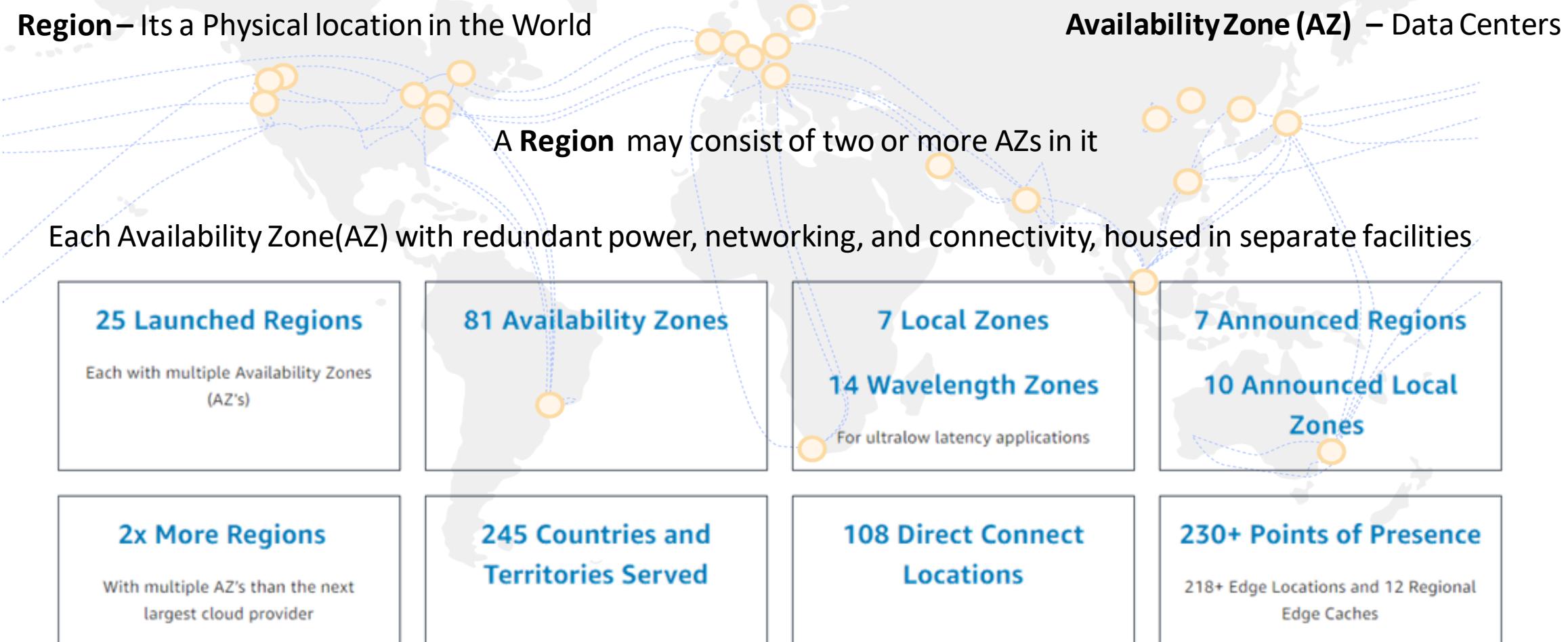


AWS Global Infrastructure



AWS now spans 81 Availability Zones within 25 geographic regions around the world, and has announced plans for 7 Regions and 10 Local Zones

Region – Its a Physical location in the World



Availability Zone (AZ) – Data Centers



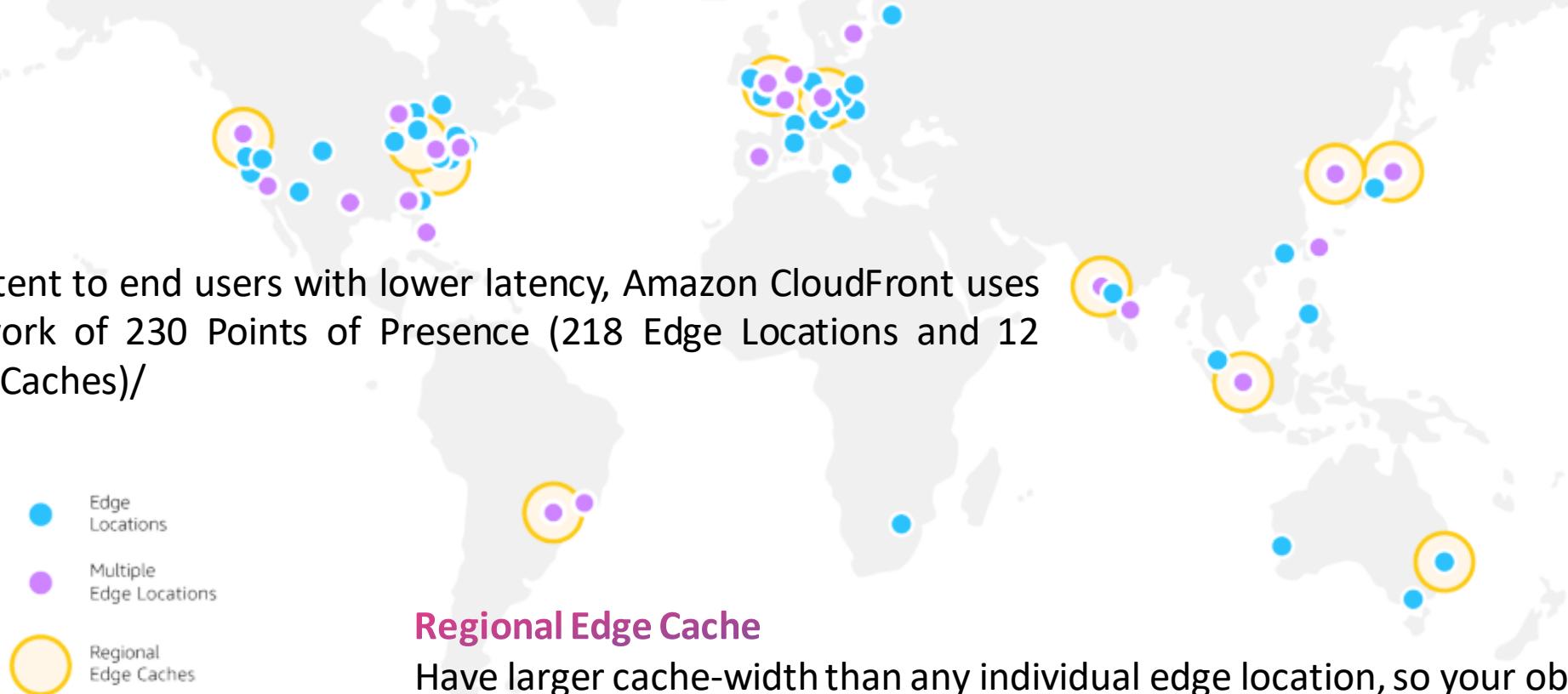
AWS Edge Locations and Regional Edge Caches

https://aws.amazon.com/about-aws/global-infrastructure/?pg=WICC-N&tile=learn_more

<https://pages.awscloud.com/GLOBAL-multi-DL-gartner-mq-cips-2020-learn.html>

Edge location

Point-of-presence (POP) location with caching servers (for Content Delivery Network)

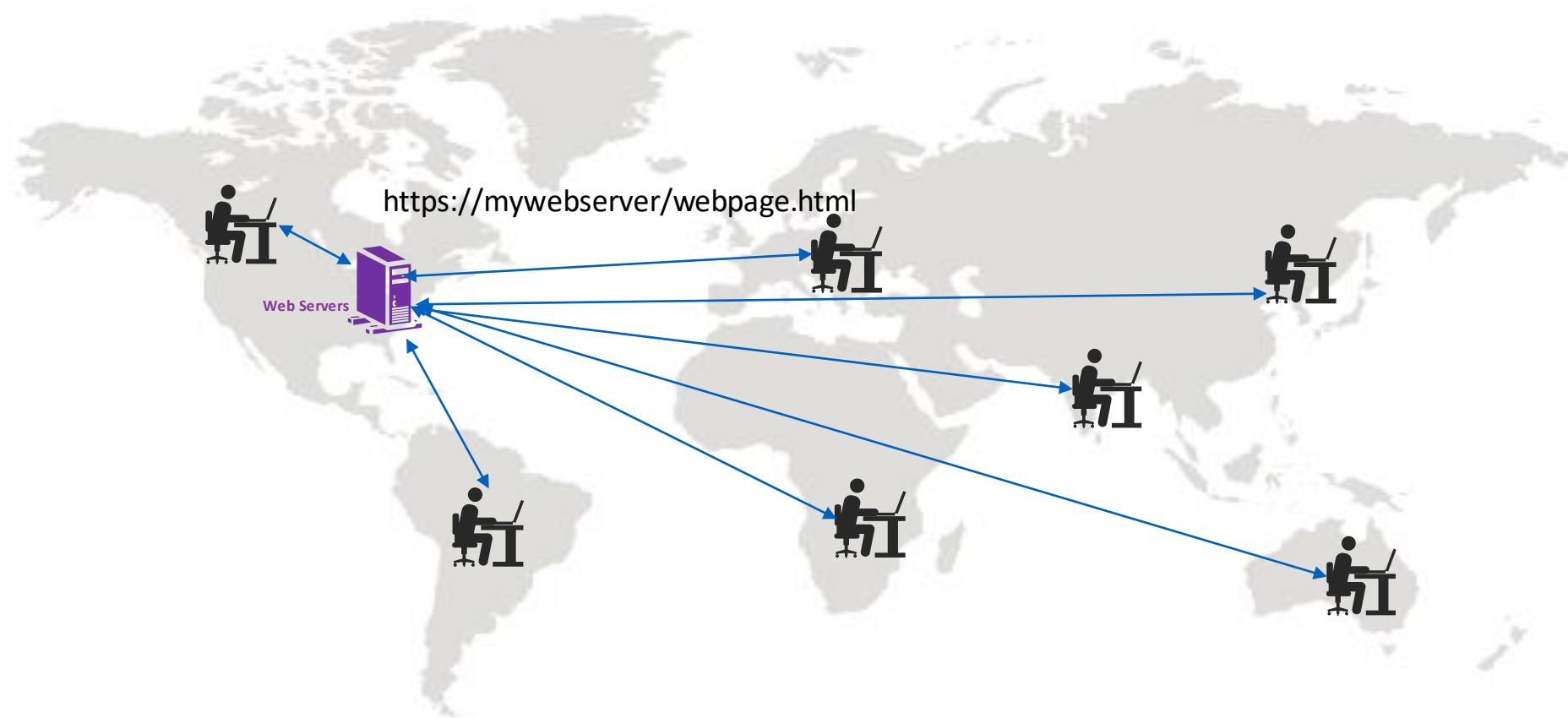


To deliver content to end users with lower latency, Amazon CloudFront uses a global network of 230 Points of Presence (218 Edge Locations and 12 Regional Edge Caches)/

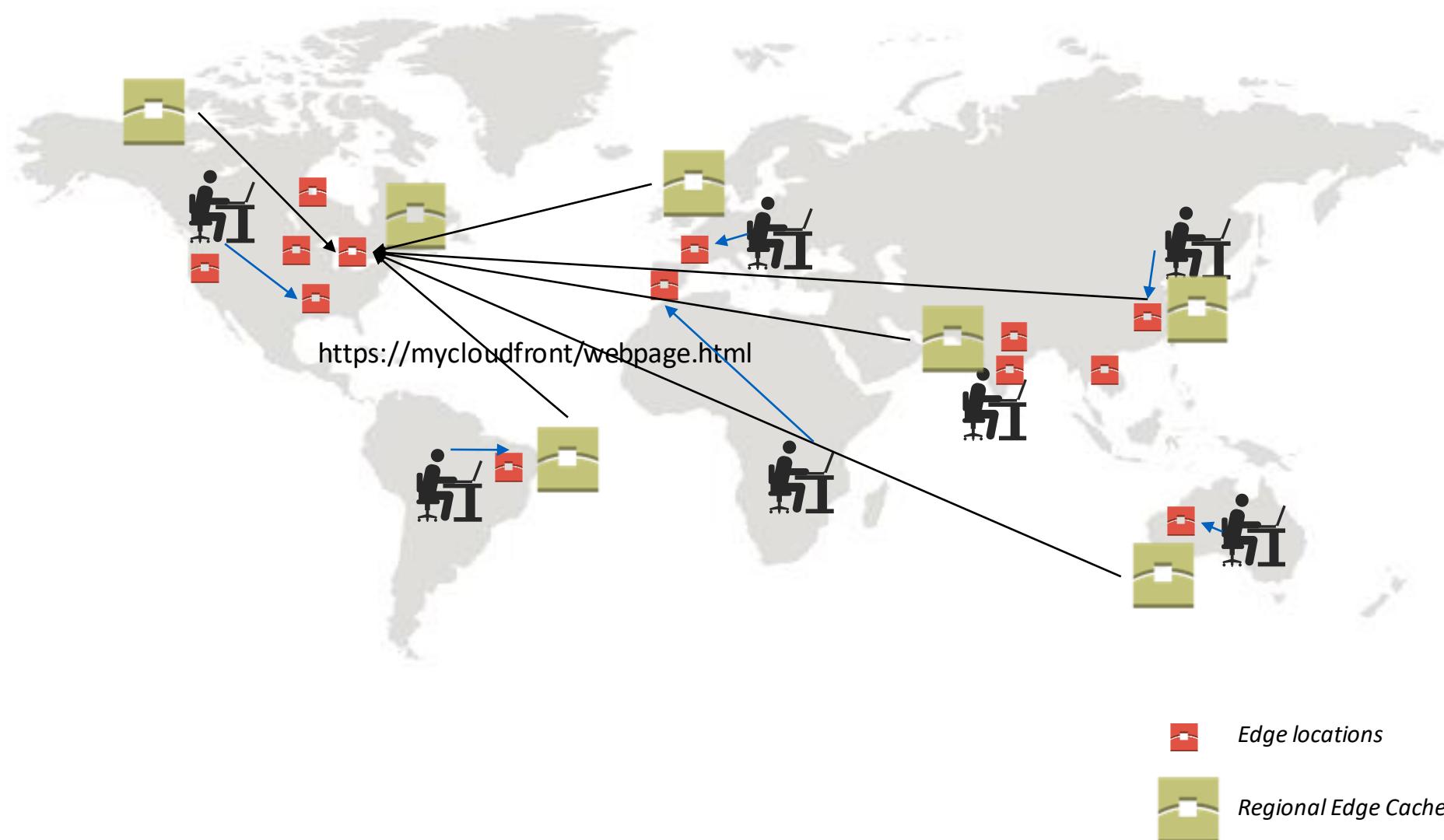
Regional Edge Cache

Have larger cache-width than any individual edge location, so your objects remain in cache longer at these locations.

AWS Edge Locations and Regional Edge Caches



AWS Edge Locations and Regional Edge Caches



AWS Edge Location

An **Edge location** is a Point-of-presence (POP) location where Amazon hosts few of its services that would provide low latency access to the end users or security to their web applications

Services offered at AWS Edge Locations:

- Amazon CloudFront
- Amazon Route 53 for DNS
- AWS Firewall Manager
- AWS Shield for DDoS protection for applications
- AWS Web Application Firewall (WAF)

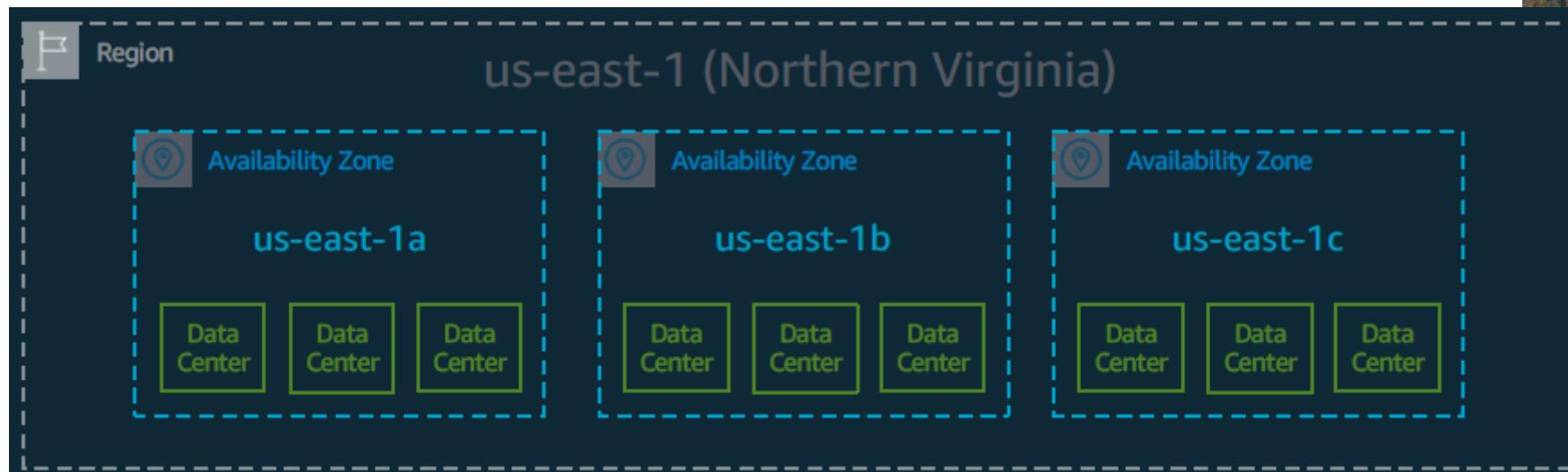


AWS Regional Edge Caches

- In Nov 2016, AWS announced addition of a new type of edge location called **Regional Edge Cache** that further improves performance for your viewers and helps reduce load on your Origin servers
- These locations sit between your origin webserver and the Global edge locations that serve traffic directly to your viewers.
- As the popularity of your objects reduce, individual edge locations may evict those objects to make room for more popular content.
- Regional Edge Caches have larger cache-width than any individual edge location, so your objects remain in cache longer at these locations.
- This helps keep more of your content closer to your viewers, reducing the need for CloudFront to go back to your origin webserver, and improving overall performance for viewers.

Availability Zones

- A region is comprised of multiple Availability Zones (typically 3)
- Fully independent partitions on isolated fault lines, flood plains, and power grids
- Each AZ: redundant power and redundant dedicated network
- Each AZ: typically multiple data centers
- Between AZs: high throughput, low latency (<10ms) network
- Between AZs: physical separation < 100km (60mi)



AWS Services Overview



Services Catalog

Compute



Storage



Database



Networking & Content Delivery



Developer Tools



Robotics



Customer Enablement



Blockchain



Migration & Transfer



Satellite



Management & Governance



Media Services



Machine Learning



Analytics



Security, Identity, & Compliance



Mobile



AR & VR



AWS Cost Management



Application Integration



Customer Engagement



Business Applications



End User Computing



Internet Of Things

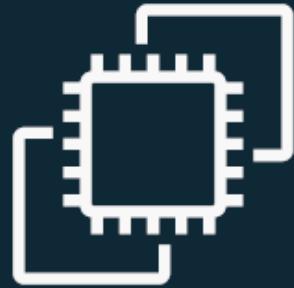


Game Development



AWS Compute

Choices for Compute



Amazon EC2

Virtual server instances
in the cloud



Amazon ECS, EKS, and Fargate

Container management
service for running
Docker on a managed
cluster of EC2



AWS Lambda

Serverless compute
for stateless code execution
in response to triggers

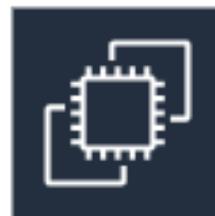


Objective

- ➔ Understand Amazon EC2
- ➔ Supported Operating Systems
- ➔ EC2 instances types
- ➔ EC2 Tenancy Options
- ➔ EC2 Purchasing options
- ➔ EC2 Instance billing
- ➔ Elastic Network Interface (ENI) and Instance IP Address
- ➔ EC2 – Amazon Machine Image (AMI)

What is EC2?

ELASTIC COMPUTE CLOUD

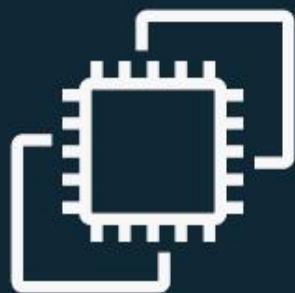


Amazon EC2

- ✓ Application Server
- ✓ Web Server
- ✓ Database Server
- ✓ Game Server
- ✓ Mail Server
- ✓ Media Server
- ✓ Catalog Server
- ✓ File Server
- ✓ Computing Server
- ✓ Proxy Server
- ✓ Etc.,

- Amazon Elastic Compute Cloud (Amazon EC2) provides scalable, resizable computing capacity (Virtual Machine) in the AWS cloud
- Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly deploy your applications, scale capacity as your computing requirements change and pay only for capacity that you actually use.
- Secure and reliable with SLA commitment of 99.95%
- Supports a variety of Linux and Windows Operating systems.
- You have full control of your instance in AWS and can manage access to it

Amazon EC2



Amazon EC2

Linux | Windows

Arm and x86 architectures

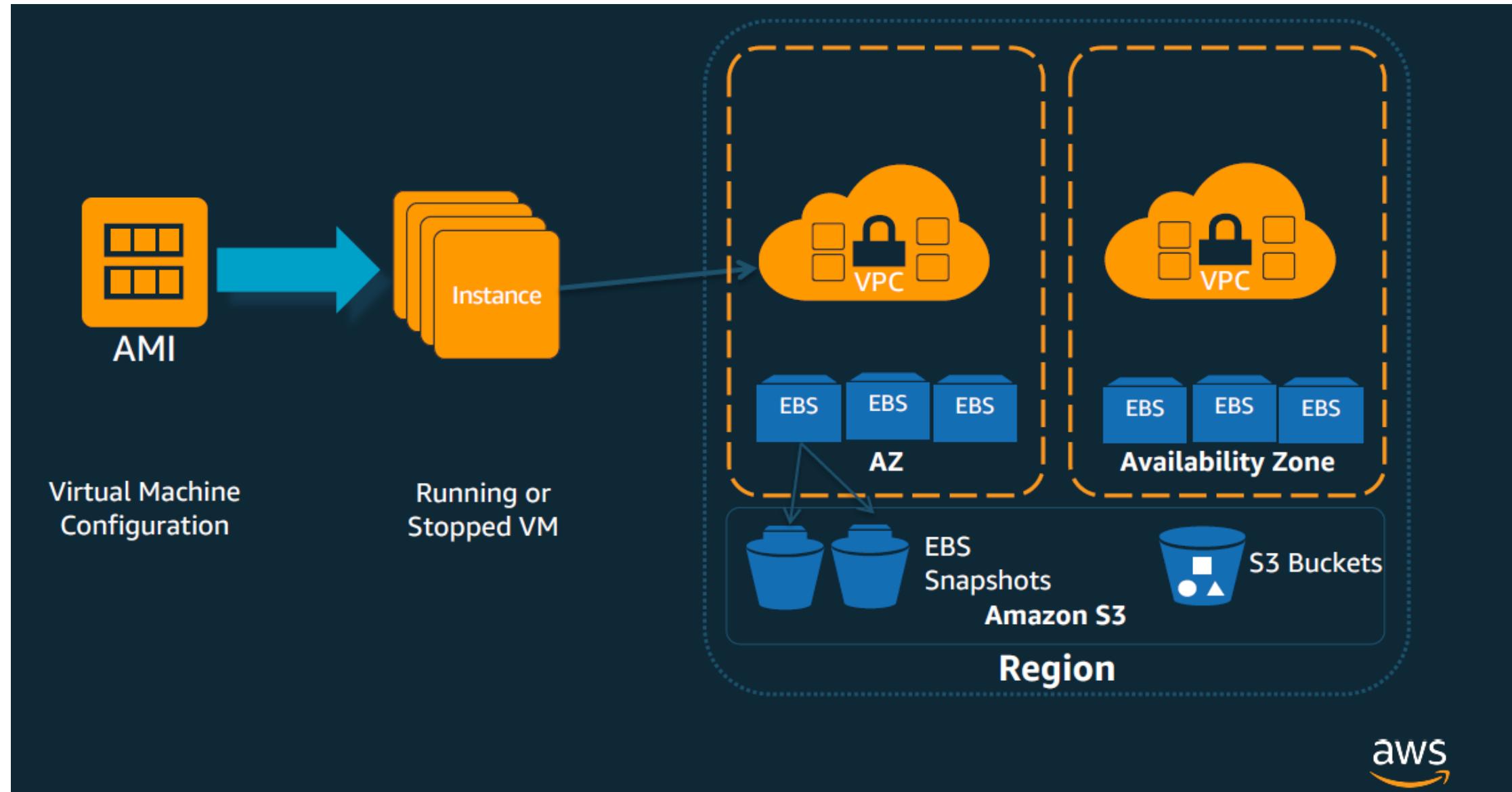
General purpose and workload optimized

Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-demand, RI, Spot

EC2 Terminology

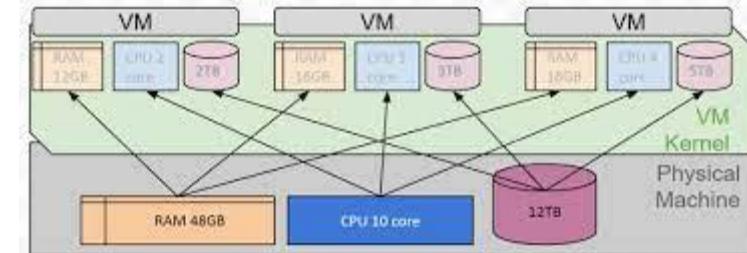
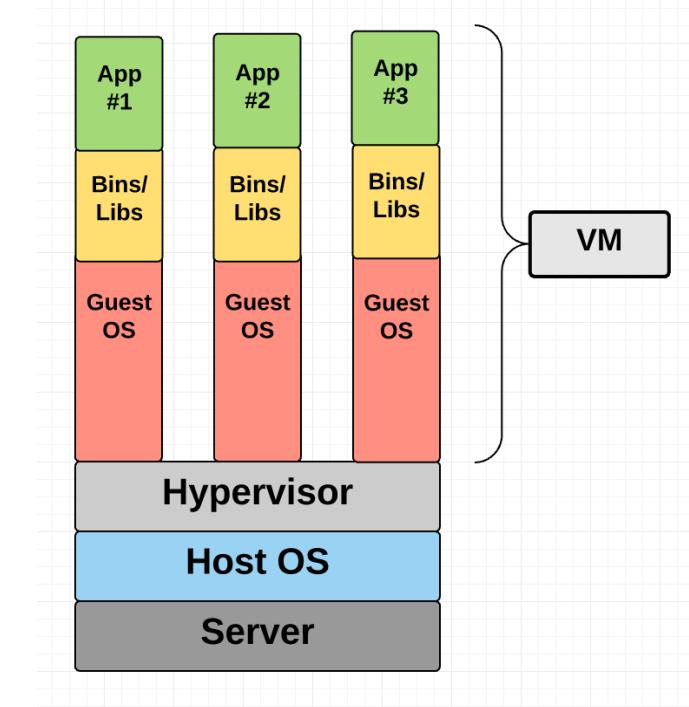


Resource allocation

- All resources assigned to you are dedicated to your instance with no over commitment
- All vCPUs are dedicated to you
- Memory allocated is assigned only to your instance
- Network resources are partitioned to avoid “noisy neighbors”

What's a virtual CPU? (vCPU)

- A vCPU is typically a hyper-threaded physical core*
- Divide vCPU count by 2 to get core count



List of supported Operating Systems



1 to 60 of 3,181 AMIs		
	amzn2-ami-hvm-2.0.20191024.3-x86_64-gp2 - ami-00dc79254d0461090 Amazon Linux 2 AMI 2.0.20191024.3 x86_64 HVM gp2 Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)
	amzn-ami-hvm-2018.03.0.20190826-x86_64-gp2 - ami-00eb20669e0990cb4 Amazon Linux AMI 2018.03.0.20190826 x86_64 HVM gp2 Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)
	amzn-ami-hvm-2018.03.0.20190826-x86_64-ebs - ami-001efa431f19d816f Amazon Linux AMI 2018.03.0.20190826 x86_64 HVM ebs Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)



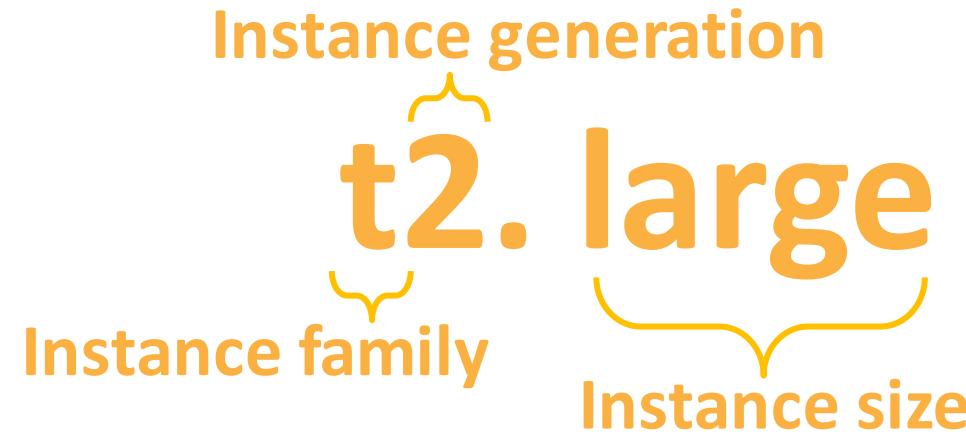
	Windows_Server-2019-English-Full-SQL_2017_Standard-2019.10.09 - ami-00da65e71856d58bb Microsoft Windows Server 2019 Full Locale English with SQL Standard 2017 AMI provided by Amazon Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)
	Windows_Server-2019-English-Full-SQL_2016_SP2_Standard-2019.10.09 - ami-099943d22372dbe2c Microsoft Windows Server 2019 Full Locale English with SQL Standard 2016 AMI provided by Amazon Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)



	EC2 CentOS 5.5 GPU HVM AMI - ami-aa30c7c3 EC2 CentOS 5.5 GPU HVM AMI Root device type: ebs Virtualization type: hvm ENA Enabled: No	<button>Select</button> 64-bit (x86)
	ultraserve-centos-6.9-ami-reverse_proxy-hvm-2019.07.5-14-x86_64-gp2 - ami-0004558ff67e36bb8 UltraServe CentOS 6.9 AMI REVERSE_PROXY - 2019.07.5-14 x86_64 HVM GP2 Root device type: ebs Virtualization type: hvm ENA Enabled: Yes	<button>Select</button> 64-bit (x86)



EC2 Instance Types



- Instance type = Instance family + instance generation
 - Choose instance type based on type /nature of application
 - Defines storage type that can be associated for additional volumes – instance store or EBS only
 - EBS optimized or NOT

Instance size = No of vCPUs, Memory, Network Bandwidth, EBS storage throughput

- Choose instance size based on resource requirement by the application

EC2 Instance Types

Instance types comprise varying ratios of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications.

<http://aws.amazon.com/ec2/instance-types/>

Instance Family	Current Generation Instance Types
General purpose	a1.medium a1.large a1.xlarge a1.2xlarge a1.4xlarge a1.metal m4.large m4.xlarge m4.2xlarge m4.4xlarge m4.10xlarge m4.16xlarge m5.large m5.xlarge m5.2xlarge m5.4xlarge m5.8xlarge m5.12xlarge m5.16xlarge m5.24xlarge m5.metal m5a.large m5a.xlarge m5a.2xlarge m5a.4xlarge m5a.8xlarge m5a.12xlarge m5a.16xlarge m5a.24xlarge m5ad.large m5ad.xlarge m5ad.2xlarge m5ad.4xlarge m5ad.12xlarge m5ad.24xlarge m5d.large m5d.xlarge m5d.2xlarge m5d.4xlarge m5d.8xlarge m5d.12xlarge m5d.16xlarge m5d.24xlarge m5d.metal m5dn.large m5dn.xlarge m5dn.2xlarge m5dn.4xlarge m5dn.8xlarge m5dn.12xlarge m5dn.16xlarge m5dn.24xlarge m5n.large m5n.xlarge m5n.2xlarge m5n.4xlarge m5n.8xlarge m5n.12xlarge m5n.16xlarge m5n.24xlarge t2.nano t2.micro t2.small t2.medium t2.large t2.xlarge t2.2xlarge t3.nano t3.micro t3.small t3.medium t3.large t3.xlarge t3.2xlarge t3a.nano t3a.micro t3a.small t3a.medium t3a.large t3a.xlarge t3a.2xlarge

EC2 Instance Types

Instance Family	Current Generation Instance Types
Compute optimized	c4.large c4.xlarge c4.2xlarge c4.4xlarge c4.8xlarge c5.large c5.xlarge c5.2xlarge c5.4xlarge c5.9xlarge c5.12xlarge c5.18xlarge c5.24xlarge c5.metal c5d.large c5d.xlarge c5d.2xlarge c5d.4xlarge c5d.9xlarge c5d.12xlarge c5d.18xlarge c5d.24xlarge c5d.metal c5n.large c5n.xlarge c5n.2xlarge c5n.4xlarge c5n.9xlarge c5n.18xlarge c5n.metal
Memory optimized	r4.large r4.xlarge r4.2xlarge r4.4xlarge r4.8xlarge r4.16xlarge r5.large r5.xlarge r5.2xlarge r5.4xlarge r5.8xlarge r5.12xlarge r5.16xlarge r5.24xlarge r5.metal r5a.large r5a.xlarge r5a.2xlarge r5a.4xlarge r5a.8xlarge r5a.12xlarge r5a.16xlarge r5a.24xlarge r5ad.large r5ad.xlarge r5ad.2xlarge r5ad.4xlarge r5ad.12xlarge r5ad.24xlarge r5d.large r5d.xlarge r5d.2xlarge r5d.4xlarge r5d.8xlarge r5d.12xlarge r5d.16xlarge r5d.24xlarge r5d.metal r5dn.large r5dn.xlarge r5dn.2xlarge r5dn.4xlarge r5dn.8xlarge r5dn.12xlarge r5dn.16xlarge r5dn.24xlarge r5n.large r5n.xlarge r5n.2xlarge r5n.4xlarge r5n.8xlarge r5n.12xlarge r5n.16xlarge r5n.24xlarge u-6tb1.metal u-9tb1.metal u-12tb1.metal u-18tb1.metal u-24tb1.metal x1.16xlarge x1.32xlarge x1e.xlarge x1e.2xlarge x1e.4xlarge x1e.8xlarge x1e.16xlarge x1e.32xlarge z1d.large z1d.xlarge z1d.2xlarge z1d.3xlarge z1d.6xlarge z1d.12xlarge z1d.metal

EC2 Instance Types

Instance Family	Current Generation Instance Types
Storage optimized	d2.xlarge d2.2xlarge d2.4xlarge d2.8xlarge h1.2xlarge h1.4xlarge h1.8xlarge h1.16xlarge i3.large i3.xlarge i3.2xlarge i3.4xlarge i3.8xlarge i3.16xlarge i3.metal i3en.large i3en.xlarge i3en.2xlarge i3en.3xlarge i3en.6xlarge i3en.12xlarge i3en.24xlarge i3en.metal
Accelerated computing	f1.2xlarge f1.4xlarge f1.16xlarge g3s.xlarge g3.4xlarge g3.8xlarge g3.16xlarge g4dn.xlarge g4dn.2xlarge g4dn.4xlarge g4dn.8xlarge g4dn.12xlarge g4dn.16xlarge p2.xlarge p2.8xlarge p2.16xlarge p3.2xlarge p3.8xlarge p3.16xlarge p3dn.24xlarge

EC2 Instance Types

Instance Family	Speciality	Use Case
D2	Dense Storage	File servers/Data Warehousing/Hadoop
R4	Memory optimized	Memory Intensive Apps/DBs
M4	General Purpose	Application Servers
C4	Compute optimized	CPU Intensive Apps/DBs
G2	Graphics Intensive	Video Encoding/3D Application Streaming
I2	High Speed Storage	NoSQL DBs, Data Warehousing etc.
F1	Field Programmable Gate Array	Hardware acceleration for your code
T2	Lowest cost, General Purpose	Web servers/small DBs
P2	Graphics/General Purpose GPU	Machine learning, Bit Coin mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc

EC2 Instance Types | Analogy

Dr. McGift PX



D	<i>Density</i>
r	<i>RAM</i>
M	<i>Main choice for General Purpose Apps</i>
c	<i>Compute</i>
G	<i>Graphics</i>
i	<i>IOPS</i>
f	<i>FPGA</i>
t	<i>Cheap general purpose (like t2 micro)</i>
P	<i>Graphics (think pixel)</i>
X	<i>Extreme Memory</i>

EC2 – Tenancy Attributes

Each instance that you launch into a VPC has a tenancy attribute. This attribute has the following values.

Default

Your instance runs on shared hardware

Dedicated

Your instance runs on single-tenant hardware.

Host

Your instance runs on a Dedicated Host, which is an isolated server with configurations that you can control.

After you launch an instance, there are some limitations to changing its tenancy.

- You cannot change the tenancy of an instance from ***default*** to ***dedicated*** or ***host*** after you've launched it.
- You cannot change the tenancy of an instance from ***dedicated*** or ***host*** to ***default*** after you've launched it.

You can change the tenancy of an instance from ***dedicated*** to ***host***, or from ***host*** to ***dedicated*** after you've launched it.

EC2 – Tenancy Attributes | VPC

Each VPC has a related instance tenancy attribute. This attribute has the following values.

Default

An instance launched into the VPC runs on shared hardware by default, unless you explicitly specify a different tenancy during instance launch.

Dedicated

An instance launched into the VPC is a Dedicated Instance by default, unless you explicitly specify a tenancy of host during instance launch. You cannot specify a tenancy of default during instance launch.

You can change the instance tenancy of a VPC from **dedicated** to **default** after you create it. You cannot change the instance tenancy of a VPC to **dedicated**.

To create Dedicated Instances, you can do the following:

- Create the VPC with the instance tenancy set to dedicated (all instances launched into this VPC are Dedicated Instances).
- Create the VPC with the instance tenancy set to default, and specify a tenancy of dedicated for any instances when you launch them.

Instance Purchasing Options

Amazon EC2 provides the following purchasing options to enable you to optimize your costs based on your needs:

On-Demand

Pay, by the second, for the instances that you launch

Savings Plans

Reduce your Amazon EC2 costs by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.

Reserved

Reduce your Amazon EC2 costs by making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years.

Scheduled

Purchase instances that are always available on the specified recurring schedule, for a one-year term

Instance Purchasing Options | Contd..

Spot

Request unused EC2 instances, which can reduce your Amazon EC2 costs significantly.

Dedicated Hosts

Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs

Dedicated Instance

Pay, by the hour, for instances that run on single-tenant hardware.

Capacity Reservations

Reserve capacity for your EC2 instances in a specific Availability Zone for any duration.

Instance Purchasing Options | On Demand

- Pay as you go model
 - Hourly charges with no upfront payment
- No need to reserve the capacity or make a long-term commitment
- Amazon Recommends On Demand instances for:
 - Users who want low cost and flexibility
 - Applications with short term, spiky, or unpredictable workloads
 - Applications being developed or tested on Amazon EC2 for the first time



Instance Purchasing Options | Reserved

- With Reserved Instances, you pay for the entire term regardless of whether you use it.
- Reservations available for 1 or 3 years
- Instance will always be available for you in the Availability Zone where it was purchased [for the purchased Operating system, Availability Zone, Instance size and Tenancy].
- Payment Options
 - No Upfront only hourly charges
 - Partial Upfront and hourly charges
 - Full upfront payment and NO hourly charges
- Reserved instances provide significant discounts, Upto75% in some cases
- Reserved instance purchase is now available for Dedicated tenancy also



Instance Purchasing Options | Scheduled

- Scheduled instances enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term
- You reserve the capacity in advance, so that you know it is available when you need it.
- You pay for the time that the instances are scheduled, even if you do not use them
- Scheduled Instances are a good choice for workloads that do not run continuously, but do run on a regular schedule
- These cost lesser than On-demand instances but you get assured compute capacity
- Scheduled Instances are subject to the following limits:
 - The following are the only supported instance types: C3, C4, M4, and R3
 - The required term is 365 days (one year).
 - The minimum required utilization is 1,200 hours per year.
 - You can purchase a Scheduled Instance up to three months in advance.



Instance Purchasing Options | Spot

- Amazon allows users to bid for its spare capacity of certain instance types
- Amazon sets a Spot price for instances which is a fraction of the normal on demand hourly rate
- Spot instances will help you in reducing the cost.
- The highest bid above the spot price gets the instance
 - Performance of the instance will be same as that of On Demand and Reserved instance
- If the spot price moves up higher than the price specified by the customer, the instance will be Terminated by Amazon
- Spot instance purchase is now available for Dedicated tenancy also
 - When using Spot instances you may want to check point the data at intervals to external storage like EBS volumes or onto S3
 - Applications should be capable of restartable transactions that did not complete on previously terminated Spot instance



EC2 –Instance Billing

- On Demand and Spot instances are billed by the hour
 - There is also a data transfer fee applicable when you use the EC2 instance
- Reserved instances have a monthly charge
- In the case of On Demand instance, you are charged for the hours that you use the instance
- In case of the Spot Instances also, you are charged for the hours used but in case Amazon terminates your instance, there is no charge associated with it for that hour.



EC2 –Instance Billing | Reserved Instance

- Monthly billing
- You will be billed for the entire month even if you have used the instance only partially in that month
- Amazon will apply the reserved instance bill to any instance (or combination of multiple instances created/deleted in that month) that match(es) the purchased reservation

- There is flexibility provided for Availability zone, Instance size. Reservation calculations also consider other linked accounts if there are no instances in the usage account to which it can be applied. Refer to following link for more details

- Refer:
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts-reserved-instances-application.html>



EC2 –Instance Billing | Dedicated tenancy

- For dedicated tenancy –Dedicated instances, there is
 - an hourly per instance usage fee
 - a dedicated per region fee (paid once per hour regardless of how many Dedicated Instances you're running)



- For dedicated tenancy –Dedicated Host
 - you pay hourly for each active Dedicated Host (and you are not billed for instance usage)



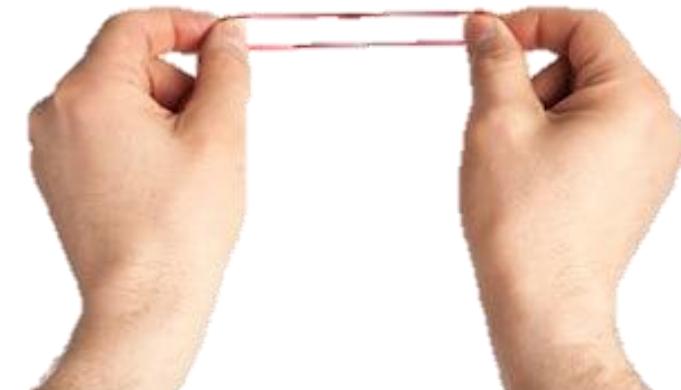
EC2 –Elastic Network Interface (ENI)

➤ Elastic Network Interface (ENI)

An elastic network interface (referred to as a network interface in this documentation) is a logical networking component in a VPC that represents a virtual network card.

A network interface can include the following attributes:

- A primary private IPv4 address from the IPv4 address range of your VPC
- One or more secondary private IPv4 addresses from the IPv4 address range of your VPC
- One Elastic IP address (IPv4) per private IPv4 address
- One public IPv4 address
- One or more IPv6 addresses
- One or more security groups
- A MAC address
- A source/destination check flag
- A description



EC2 –Instance IP Address | Private IP Address

- Private IP Address
 - Every instance you launch within AWS VPC gets a private IP address (which is allocated from the IP range used for the AWS VPC network) using either DHCP or by choosing IP manually
 - Private IP Address is used to communicate with other instances with the Amazon infrastructure
 - You can assign one primary private IP and multiple secondary IP addresses to the same instance
 - Private IP address for an instance does not change with instance start / stop and is permanent

PRIVATE

- Private DNS name(*.internal)
 - Each instance is also given an internal DNS hostname that resolves to the private IPv4 address of the instance; for example, ip-10-251-50-12.ec2.internal



EC2 –Instance IP Address | Public IP Address

- In addition to the Private IP address, you can optionally have a Public IP address (Public IP means an Internet routable IP address) assigned to your instance if you need Internet accessibility for your instance
- There are two ways to assign Public IP address to your instance
 - Auto-assigned Public IP
 - ❖ Public IP of an instance changes if you stop and start your instance
 - ❖ There is no extra charge for auto-assigned Public IP
 - Manually assigned static Public IP –called as Elastic IP address
 - ❖ You can get an Elastic IP address allocated to you within your AWS account
 - ❖ Elastic IP does not change across instance start / stop
 - ❖ You may choose to reassign the Elastic IP to a different instance (e.g. to enable application to be used from a standby instance incase of primary instance failure)
 - ❖ It remains in your account until you choose to release it and you are charged for it
- Public DNS name(*.amazonaws.com)
 - Each instance also gets a Public DNS name if a Public IP is assigned to the instance



EC2 –Instance IP Address

Public IP address automatically assigned to the instance

Instance ID	i-0a37b3f9d2e1e9db2	Public DNS (IPv4)	ec2-13-126-190-235.ap-south-1.compute.amazonaws.com
Instance state	running	IPv4 Public IP	13.126.190.235
Instance type	t2.micro	IPv6 IPs	-
Elastic IPs		Private DNS	ip-172-31-24-132.ap-south-1.compute.internal
Availability zone	ap-south-1a	Private IPs	172.31.24.132
Security groups	nin-linuxsg1. view inbound rules	Secondary private IPs	172.31.20.177

Static Public IP address (called as Elastic IP) manually assigned to the instance

Instance ID	i-0a37b3f9d2e1e9db2	Public DNS (IPv4)	ec2-13-126-22-220.ap-south-1.compute.amazonaws.com
Instance state	running	IPv4 Public IP	13.126.22.220
Instance type	t2.micro	IPv6 IPs	-
Elastic IPs	13.126.22.220*	Private DNS	ip-172-31-24-132.ap-south-1.compute.internal
Availability zone	ap-south-1a	Private IPs	172.31.24.132
Security groups	nin-linuxsg1. view inbound rules	Secondary private IPs	172.31.20.177

EC2 - Amazon Machine Images (AMI)

An Amazon Machine Image (AMI) provides the information required to launch an instance. You must specify an AMI when you launch an instance. You can launch multiple instances from a single AMI when you need multiple instances with the same configuration. You can use different AMIs to launch instances when you need instances with different configurations.



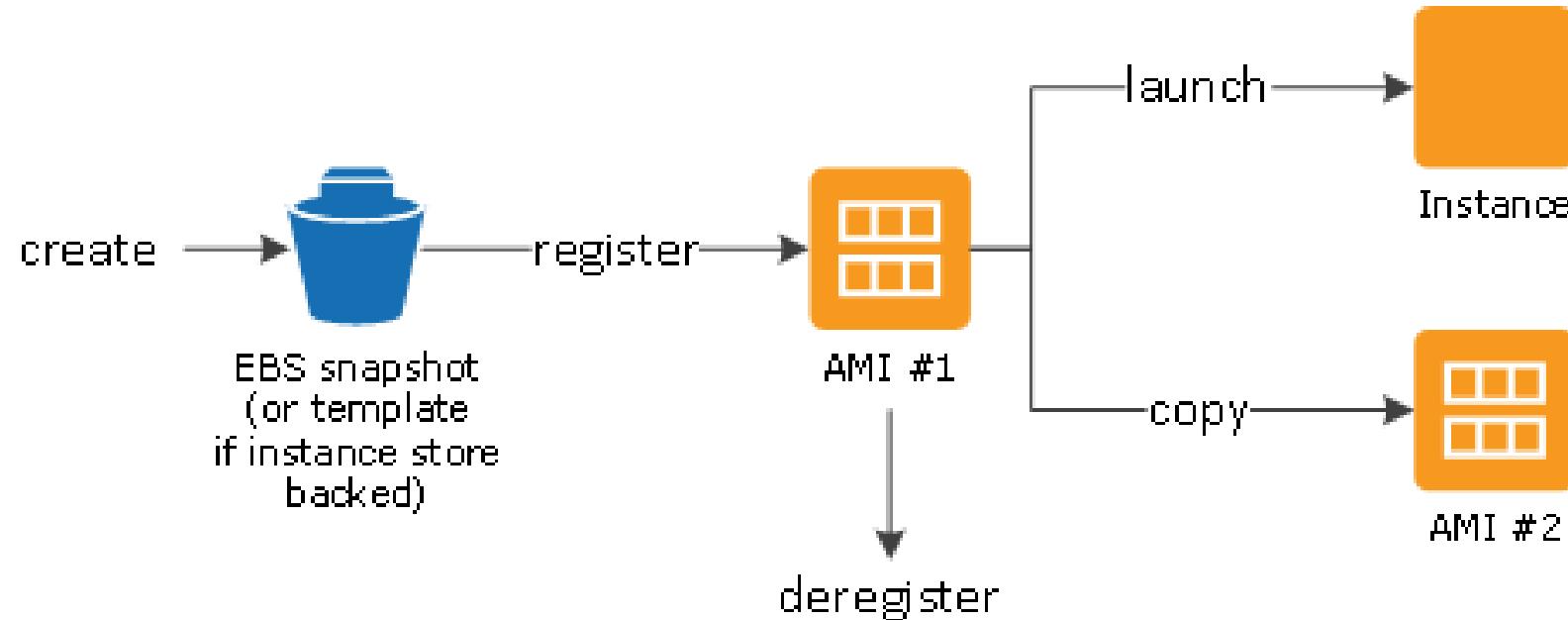
Amazon Machine
Image (AMI)

An AMI includes the following:

- One or more EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance (for example, an operating system, an application server, and applications).
- Launch permissions that control which AWS accounts can use the AMI to launch instances.
- A block device mapping that specifies the volumes to attach to the instance when it's launched.

EC2 - Amazon Machine Images (AMI) | Lifecycle

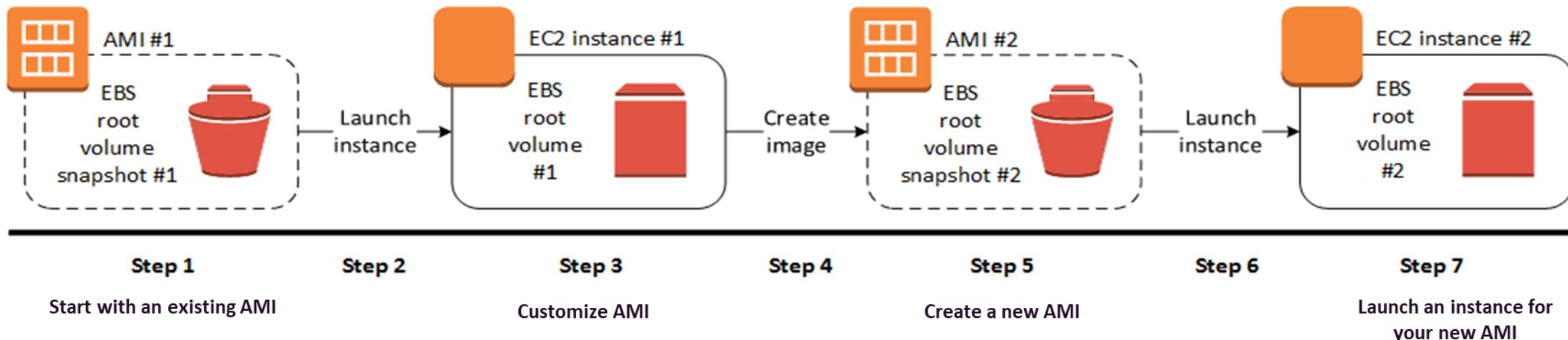
The following diagram summarizes the AMI lifecycle. After you create and register an AMI, you can use it to launch new instances. (You can also launch instances from an AMI if the AMI owner grants you launch permissions.)



You can copy an AMI within the same region or to different regions. When you no longer require an AMI, you can deregister it.

EC2 - Amazon Machine Images (AMI) | Lifecycle

You can create an AMI using the AWS Management Console or the command line. The following diagram summarizes the process for creating an Amazon EBS-backed AMI from a running EC2 instance.



Start with an existing AMI, launch an instance, customize it, create a new AMI from it, and finally launch an instance of your new AMI.

Boot volume from Amazon AMI cannot be encrypted, only additional EBS volumes can be encrypted

To have encrypted boot device volume

- Use third party software like bit locker
- Launch an instance, create an AMI image from it, create a copy of this AMI and while copying select the encryption option



Thank
You