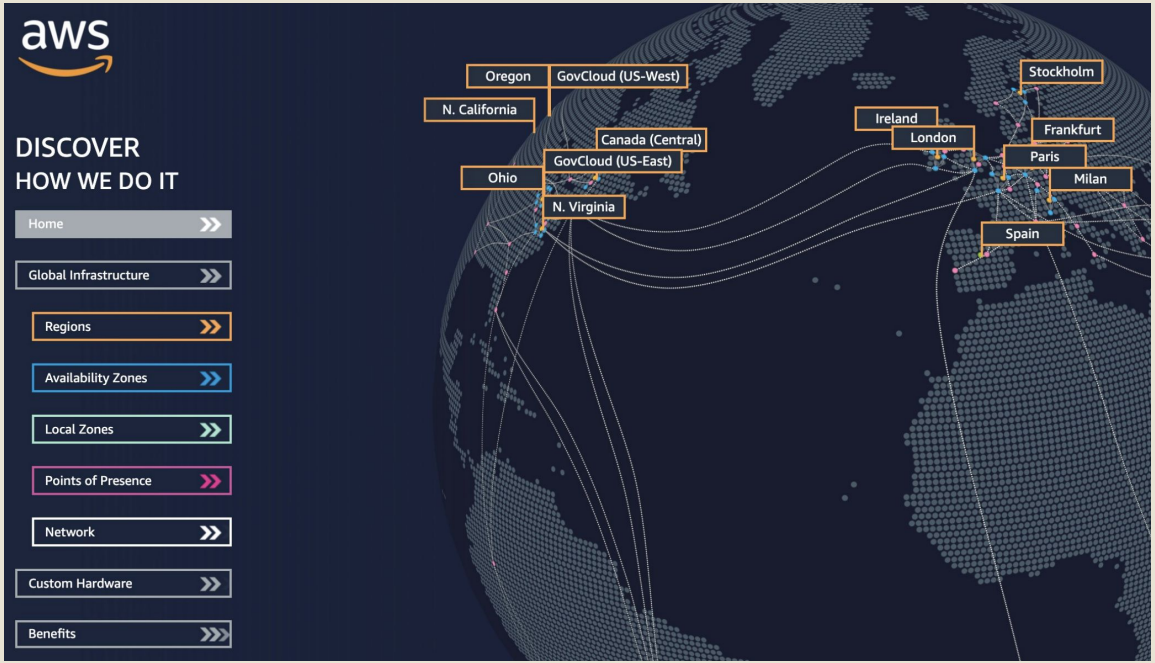
Compute: Amazon EC2 & EBS

SOMMAIRE

AWS Introduction

AWS Global Infrastructure

- AWS Regions
- AWS Availability Zones
- AWS Data Centers
- AWS Edge Locations / Points of Presence
- https://infrastructure.aws/



AWS Regions

- AWS has Regions all around the world
- Names can be us-east-1, eu-west-3…
- A region is a cluster of data centers
- Most AWS services are region-scoped

https://aws.amazon.com/about-aws/global-infrastructure/

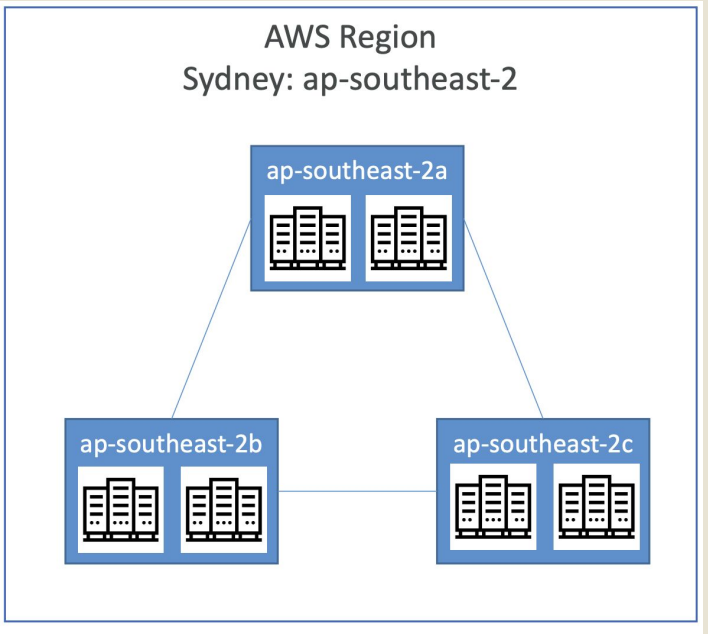
How to choose an AWS Region?

- **Compliance with data governance and legal requirements:** data never leaves a region without your explicit permission
- **Proximity to customers:** reduced latency
- **Available services within a Region:** new services and new features aren’t available in every Region
- **Pricing:** pricing varies region to region and is transparent in the service pricing page

AWS Availability Zones

- Each region has many availability zones (usually 3, min is 2, max is 6). Example:
  - ap-southeast-2a
  - ap-southeast-2b
  - ap-southeast-2c

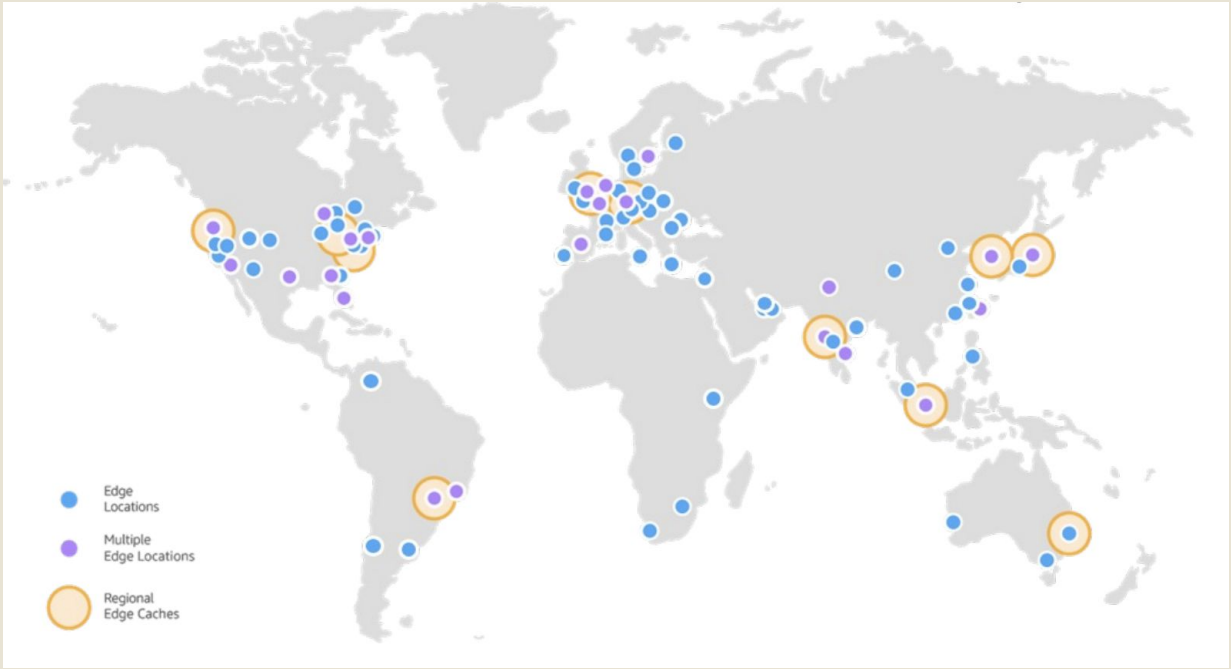
- Each availability zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity
- They’re separate from each other, so that they’re isolated from disasters
- They’re connected with high bandwidth, ultra-low latency networking



AWS Points of Presence (Edge Locations)

- At time of writing, Amazon has 216 Points of Presence (205 Edge Locations & 11 Regional Caches) in 84 cities across 42 countries.

- Content is delivered to end users with lower latency



Tour of the AWS Console

AWS has Global Services:

- Identity and Access Management (IAM)
- Route 53 (DNS service)
- CloudFront (Content Delivery Network)
- WAF (Web Application Firewall)

Most AWS services are Region-scoped:

- Amazon EC2 (Infrastructure as a Service)
- Elastic Beanstalk (Platform as a Service)
- Lambda (Function as a Service)
- Rekognition (Software as a Service)

- **Region Table:**
  https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services

EC2 Basics

Amazon EC2

- EC2 is one of the most popular of AWS offering
- EC2 = Elastic Compute Cloud = Infrastructure as a Service
- It mainly consists in the capability of :
    - Renting virtual machines (EC2)
    - Storing data on virtual drives (EBS)
    - Distributing load across machines (ELB)
    - Scaling the services using an auto-scaling group (ASG)

- Knowing EC2 is fundamental to understand how the Cloud works

EC2 sizing & configuration options

- Operating System (OS): Linux, Windows or Mac OS
- How much compute power & cores (CPU)
- How much random-access memory (RAM)
- How much storage space:

  - Network-attached (EBS & EFS)
  - hardware (EC2 Instance Store)

- Network card: speed of the card, Public IP address
- Firewall rules: security group
- Bootstrap script (configure at first launch): EC2 User Data

Amazon Elastic Compute Cloud

An EC2 instance is a virtual server

EC2 Host Server

EC2 Instance EC2 Instance EC2 Instance EC2 Instance

EC2 Instance EC2 Instance EC2 Instance EC2 Instance

EC2 Instance EC2 Instance EC2 Instance EC2 Instance

EC2 instances run Windows or Linux OS

Launching an Amazon EC2 instance

Amazon Machine Image (AMI)

Instance Type

Family	Type	vCPUs	Memory (GiB)
General purpose	t2.micro	1	1
Compute optimized	c5n.large	2	5.25
Memory optimized	r5ad.large	2	16
Storage optimized	d2.xlarge	4	30.5
GPU instances	g2.2xlarge	8	15

Amazon EC2 Reserved Instances

Burstable instances

- ✔ T3, T3a, and T2 instances, are designed to provide a baseline level of CPU performance with the ability to burst to a higher level when required.
- ✔ Burstable performance instances are the only instance types that use credits for CPU usage.
- ✔ A CPU credit provides for 100% utilization of a full CPU core for one Minute.
- ✔ Each burstable performance instance continuously earns (at a millisecond-level resolution) a set rate of CPU credits per hour, depending on the instance size.

Amazon EC2 Reserved Instances

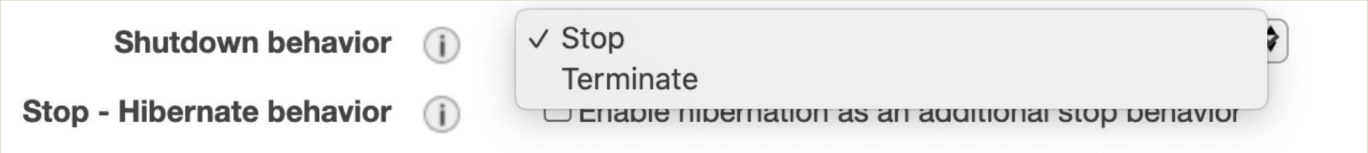
T2/T3 Unlimited:

- ✔ T2 instances are a low-cost, general purpose instance type that provides a baseline level of CPU performance with the ability to burst above the baseline when needed
- ✔ T2 Unlimited instances can sustain high CPU performance for as long as a workload needs it
- ✔ The baseline performance and ability to burst are governed by CPU Credits
- ✔ T2 instances accumulate CPU Credits when they are idle, and consume CPU Credits when they are active

Amazon EC2 Reserved Instances

Shutdown behavior

- Configure to Stop or Terminate (applies to OS-level shutdown)
  Can additionally enable hibernation (stores contents of RAM on the root volume).



Termination Protection

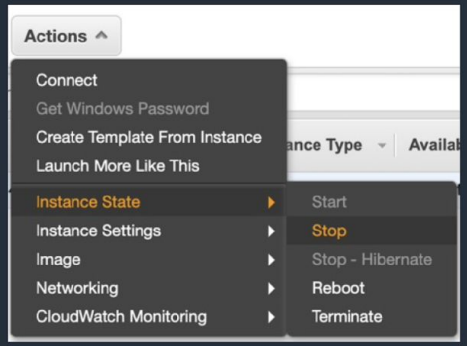
- You can protect instances from being accidentally terminated

- Once enabled, you won't be able to terminate the instance via the API or the AWS

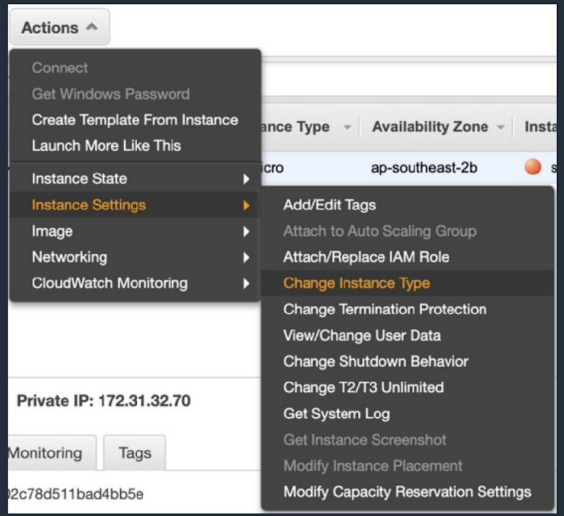
- Management Console until termination protection has been disabled

How to Change the EC2 Instance Type

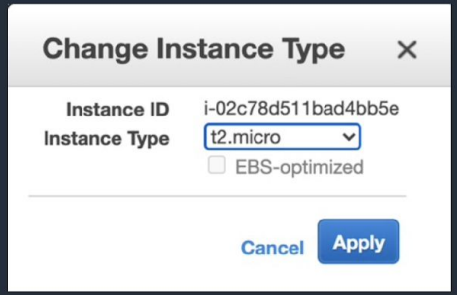
Stop the Instance



Select “Change Instance Type”



Choose the new instance type



You can change instance types for EBS backed instances only

Amazon EC2 Placement Groups

- **Cluster:**

Packs instances close together inside an Availability Zone. This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.

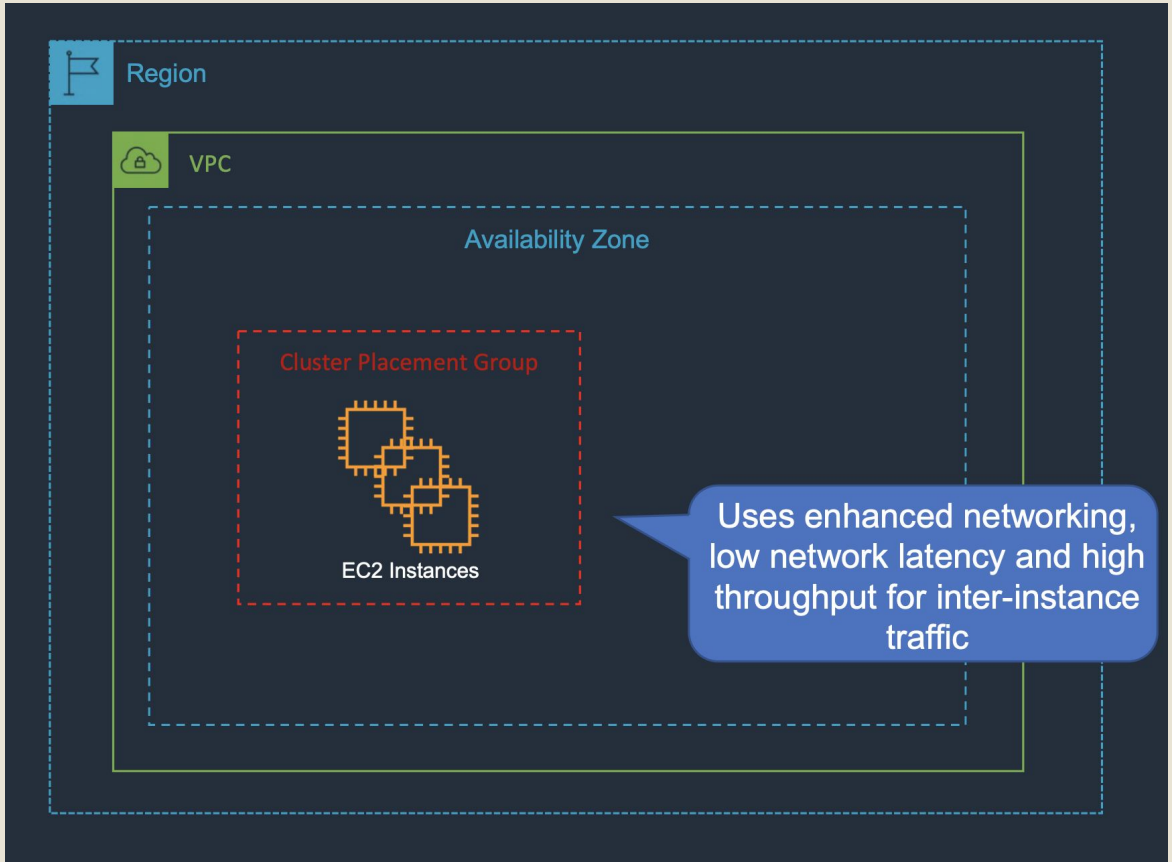
- **Partition:**

Spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.

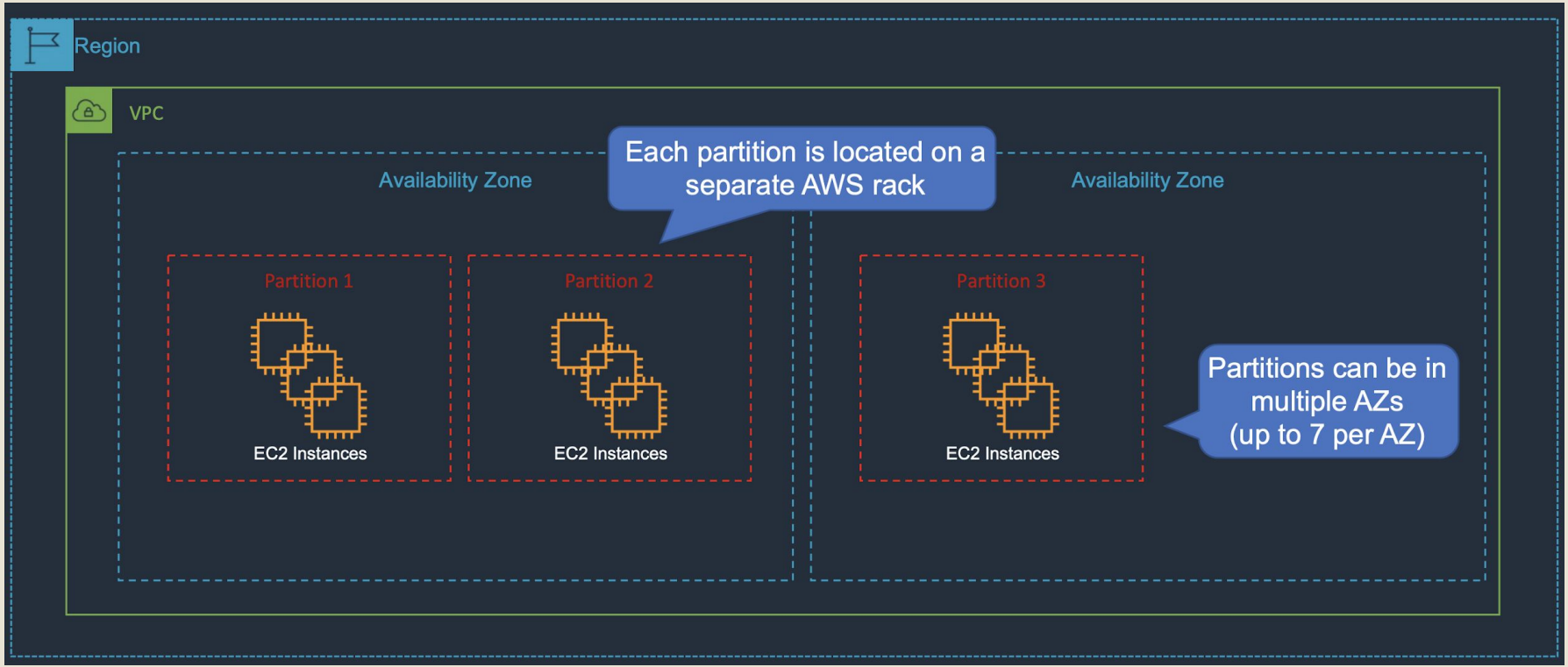
- **Spread:**

Strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

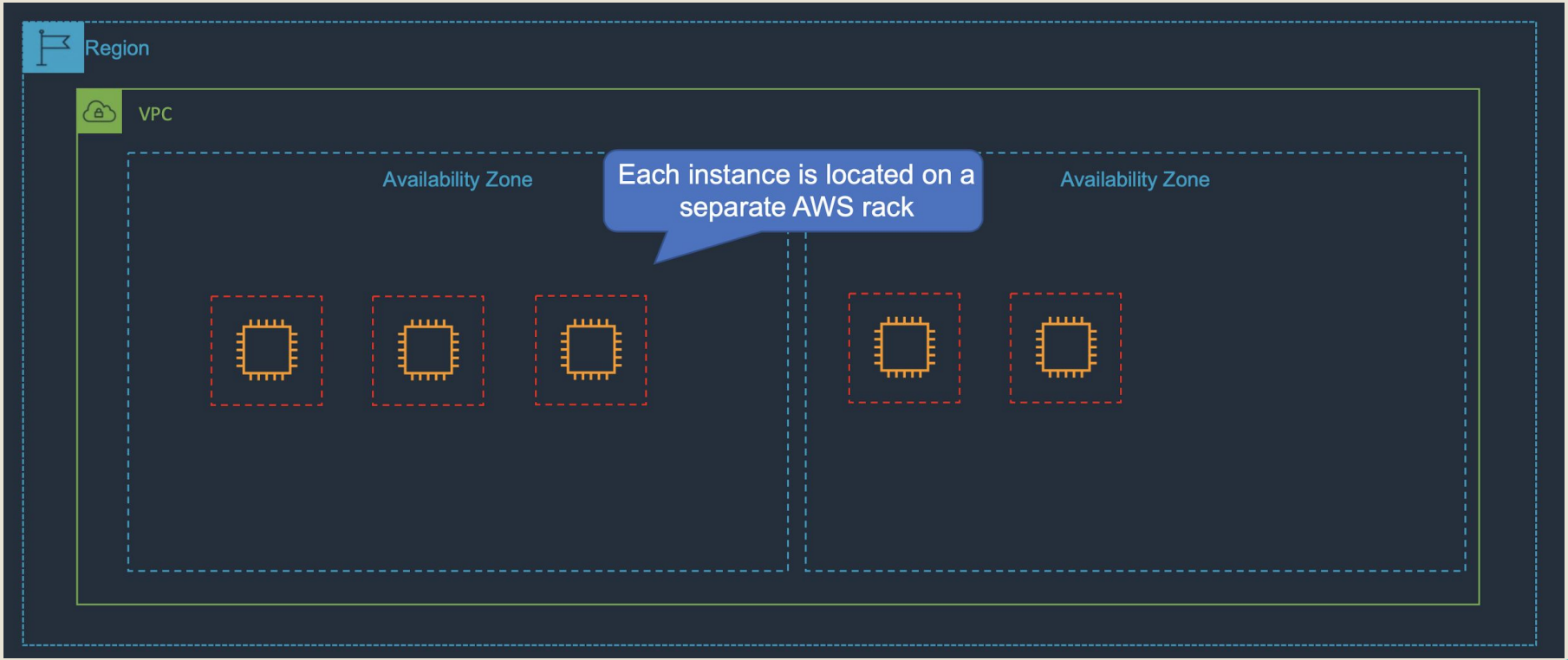
Cluster Placement Groups



Partition Placement Groups



Spread Placement Groups



Amazon EC2 Placement Groups

	Clustered	Spread	Partition
What	Instances are placed into a low-latency group within a single AZ	Instances are spread across underlying hardware	Instances are grouped into logical segments called partitions which use distinct hardware
When	Need low network latency and/or high network throughput	Reduce the risk of simultaneous instance failure if underlying hardware fails	Need control and visibility into instance placement
Pros	Get the most out of enhanced networking Instances	Can span multiple AZs	Reduces likelihood of correlated failures for large workloads.
Cons	Finite capacity: recommend launching all you might need up front	Maximum of 7 instances running per group, per AZ	Partition placement groups are not supported for Dedicated Hosts

Amazon EC2 Pricing Models

On-Demand	Reserved Instances	Savings Plans	Spot
No upfront fee	Options: No upfront, partial upfront or all upfront	Options: No upfront, partial upfront or all upfront	No upfront fee
Charged by hour or second	Charged by hour or second	Charged based on $/hour	Charged by hour or second
No commitment	1-year or 3-year commitment	1-year or 3-year commitment	No commitment
Ideal for short term needs or unpredictable workloads	Ideal for steady-state workloads and predictable usage	More flexibility: Applies across Regions and instance families/types	Ideal for cost-sensitive, compute intensive use cases that can withstand interruption

Amazon EC2 Reserved Instances

- A Reserved Instance has four instance attributes that determine its price:

- **Instance type:** For example, m4.large
- **Region:** The Region in which the Reserved Instance is purchased
- **Tenancy:** Whether your instance runs on shared (default) or single-tenant (dedicated) hardware
- **Platform:** The operating system; for example, Windows or Linux/Unix

- <u>Term commitment:</u>

- One-year: A year is defined as 31536000 seconds (365 days)

- Three-year: Three years is defined as 94608000 seconds (1095 days)

Amazon EC2 Reserved Instances

Payment Options

- **All Upfront:**

Full payment is made at the start of the term, with no other costs or additional hourly charges incurred for the remainder of the term, regardless of hours used

- **Partial Upfront:**

A portion of the cost must be paid upfront and the remaining hours in the term are billed at a discounted hourly rate, regardless of whether the Reserved Instance is being used.

- **No Upfront:**

You are billed a discounted hourly rate for every hour within the term, regardless of whether the Reserved Instance is being used

Amazon EC2 Reserved Instances

Offering class

- Standard:

These provide the most significant discount but can only be modified.

- Convertible:

These provide a lower discount than Standard Reserved Instances but can be exchanged for another Convertible Reserved Instance with different instance attributes.

Amazon EC2 Reserved Instances

Standard Reserved Instance	Convertible Reserved Instance
Some attributes, such as instance size, can be modified during the term; however, the instance family cannot be modified. You cannot exchange a Standard Reserved Instance, only modify it.	Can be exchanged during the term for another Convertible Reserved Instance with new attributes including instance family, instance type, platform, scope, or tenancy. You can also modify some attributes of a Convertible Reserved Instance.
Can be sold in the Reserved Instance Marketplace.	Cannot be sold in the Reserved Instance Marketplace.

Amazon EC2 Dedicated Instances and Hosts

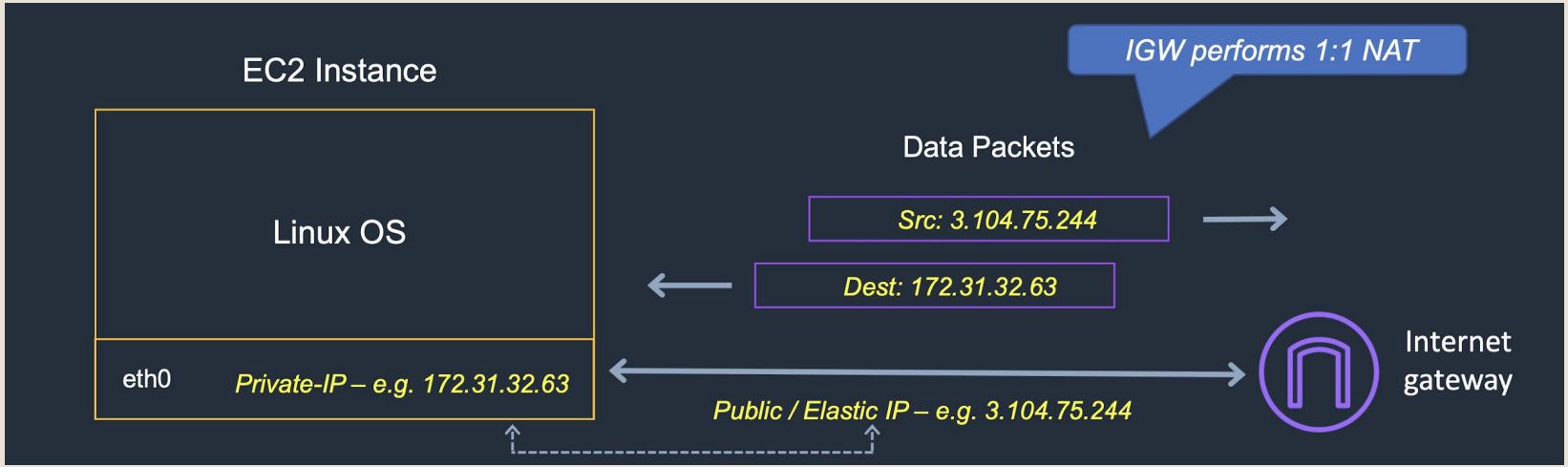
Characteristic	Dedicated Instances	Dedicated Hosts
Enables the use of dedicated physical servers	X	X
Per instance billing (subject to a $2 per region fee)	X	
Per host billing		X
Visibility of sockets, cores, host ID		X
Affinity between a host and instance		X
Targeted instance placement		X
Automatic instance placement	X	X
Add capacity using an allocation request		X

Public, Private, and Elastic IP addresses

Name	Description
Public IP address	Lost when the instance is stopped Used in Public Subnets No charge Associated with a private IP address on the instance Cannot be moved between instances
Private IP address	Retained when the instance is stopped Used in Public and Private Subnets
Elastic IP address	Static Public IP address You are charged if not used Associated with a private IP address on the instance Can be moved between instances and Elastic Network Adapters

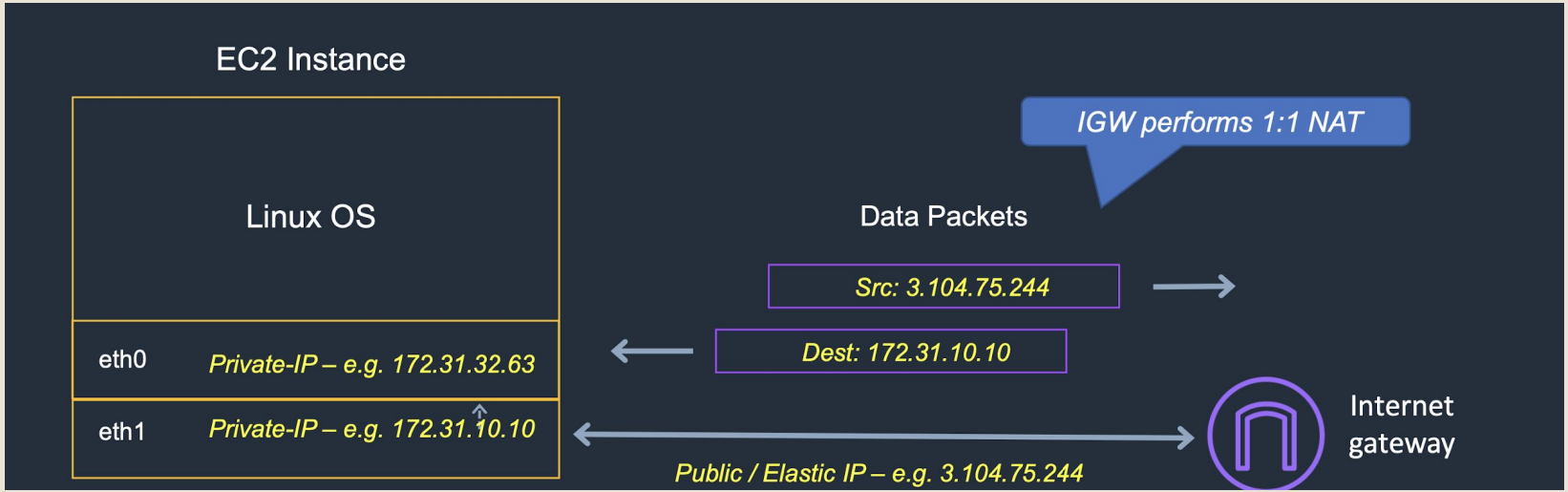
Public, Private, and Elastic IP addresses

```
[ec2-user@ip-172-31-32-63 ~]$ ip addr show eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9001 qdisc pfifo_fast sta
te UP group default qlen 1000
    link/ether 06:06:db:ad:56:28 brd ff:ff:ff:ff:ff:ff
    inet 172.31.32.63/20 brd 172.31.47.255 scope global dynamic eth0
       valid_lft 3330sec preferred_lft 3330sec
    inet6 fe80::406:dbff:fead:5628/64 scope link
       valid_lft forever preferred_lft forever
```



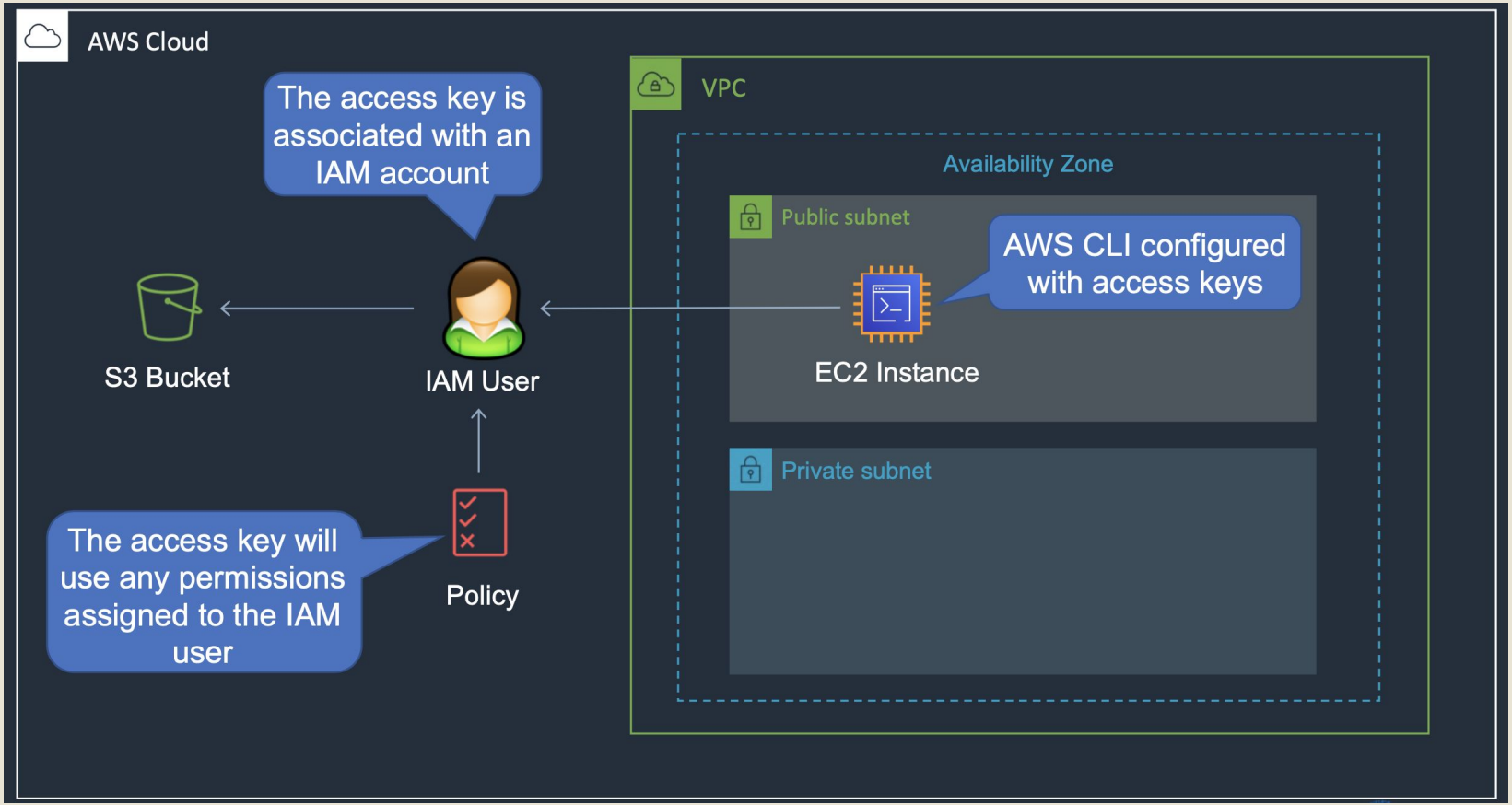
The public IP / EIP is associated with the instance

Public, Private, and Elastic IP addresses – Additional ENI

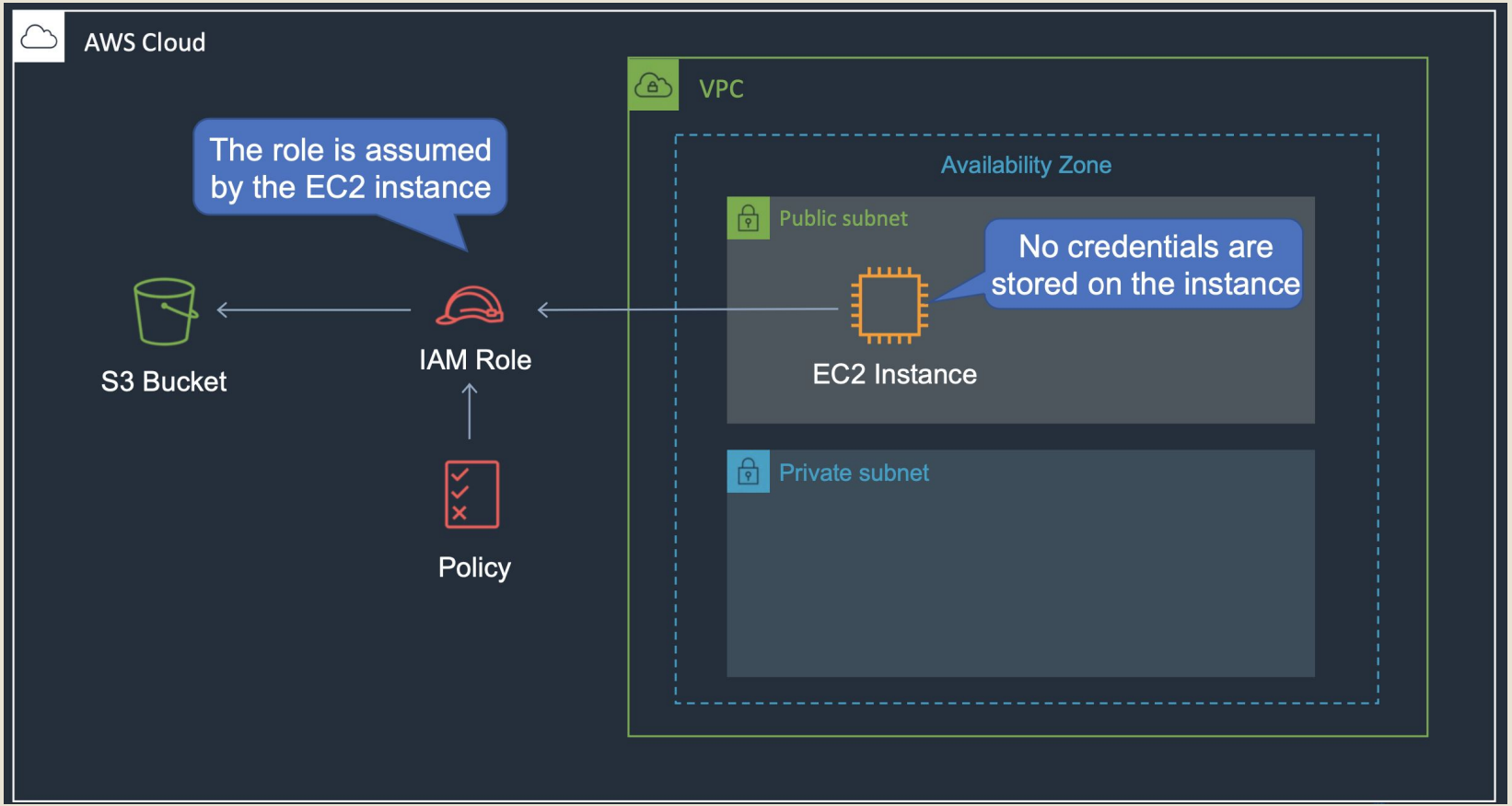


Additional Elastic Network Interface (ENI) attached

Accessing other AWS Services Using Access Keys

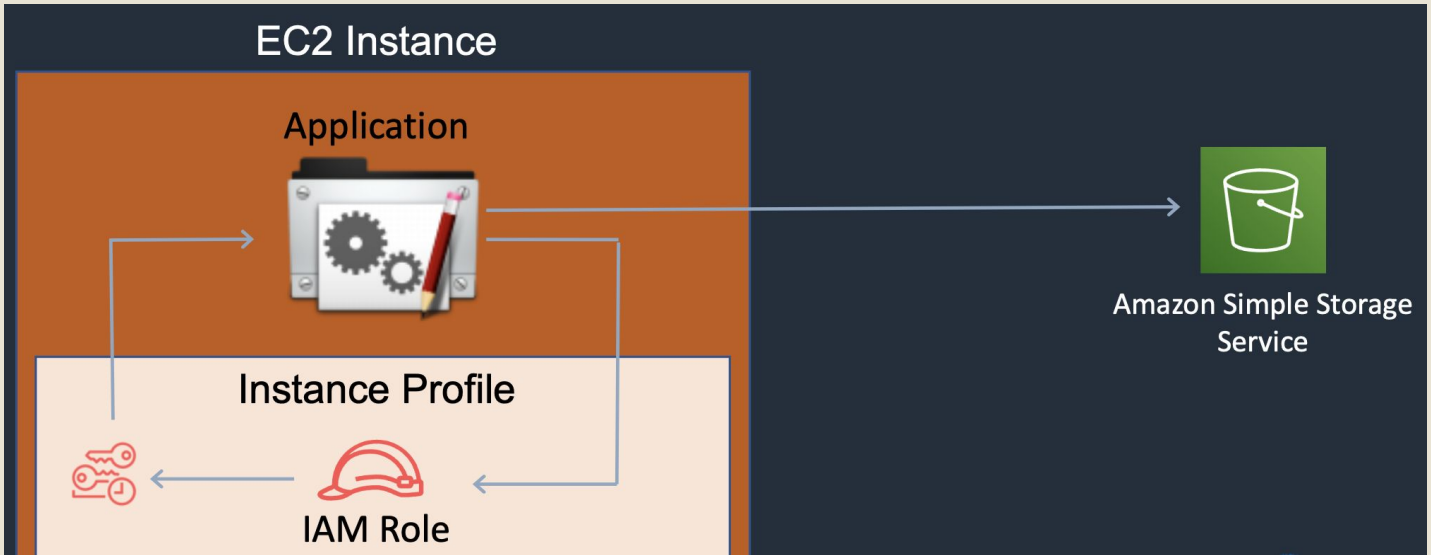


Accessing other AWS Services Using IAM Roles



IAM Instance Profiles

- ✔ An instance profile is a container for an IAM role that you can use to pass role information to an EC2 instance when the instance starts.
- ✔ An instance profile can contain only one IAM role, although a role can be included in multiple instance profiles.



IAM Instance Profiles

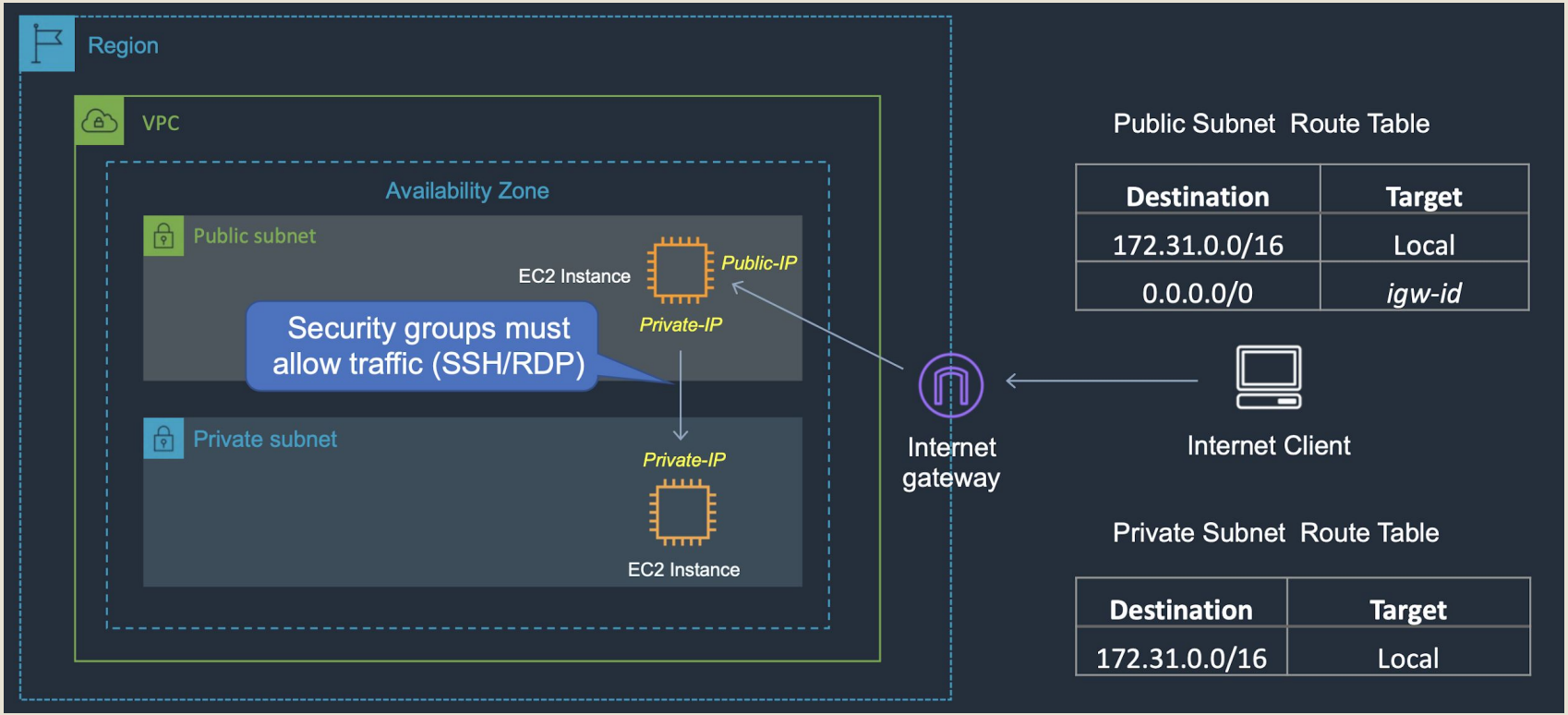
You can use the following AWS CLI commands to work with instance profiles:

- Create an instance profile: aws iam create-instance-profile
- Add a role to an instance profile: aws iam add-role-to-instance-profile
- List instance profiles: aws iam list-instance-profiles, aws iam

list-instance-profiles-for-role

- Get information about an instance profile: aws iam get-instance-profile
- Remove a role from an instance profile: aws iam remove-role-from-instance-profile
- Delete an instance profile: aws iam delete-instance-profile

Private Subnets and Bastion Hosts



Public Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*igw-id*

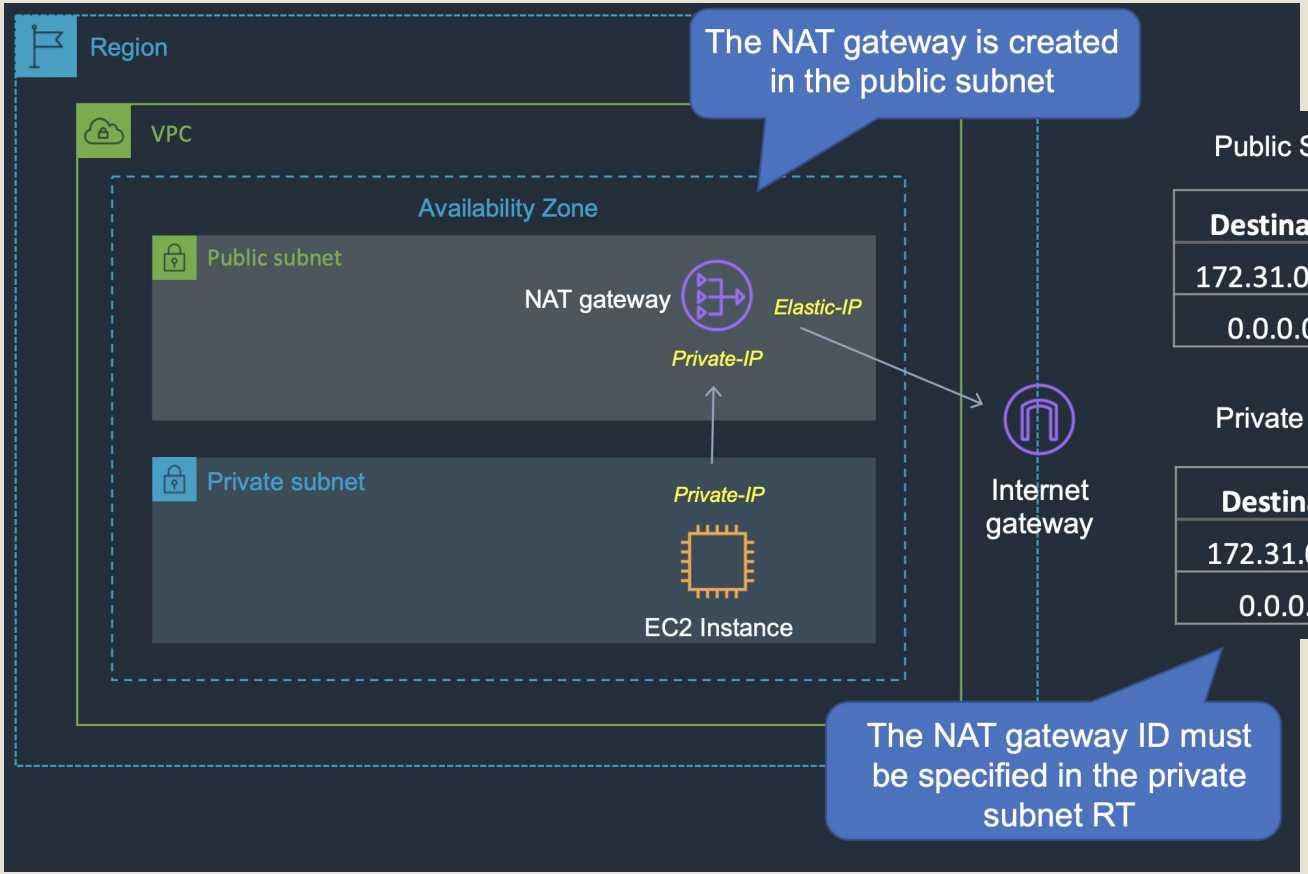
Private Subnet Route Table

Destination	Target
172.31.0.0/16	Local

NAT Instance vs NAT Gateway

NAT Instance	NAT Gateway
Managed by you (e.g. software updates)	Managed by AWS
Scale up (instance type) manually and use enhanced networking	Elastic scalability up to 45 Gbps
No high availability – scripted/auto-scaled HA possible using multiple NATs in multiple subnets	Provides automatic high availability within an AZ and can be placed in multiple AZs
Need to assign Security Group	No Security Groups
Can use as a bastion host	Cannot access through SSH
Use an Elastic IP address or a public IP address with a NAT instance	Choose the Elastic IP address to associate with a NAT gateway at creation
Can implement port forwarding through manual customisation	Does not support port forwarding

Private Subnet with NAT Gateway



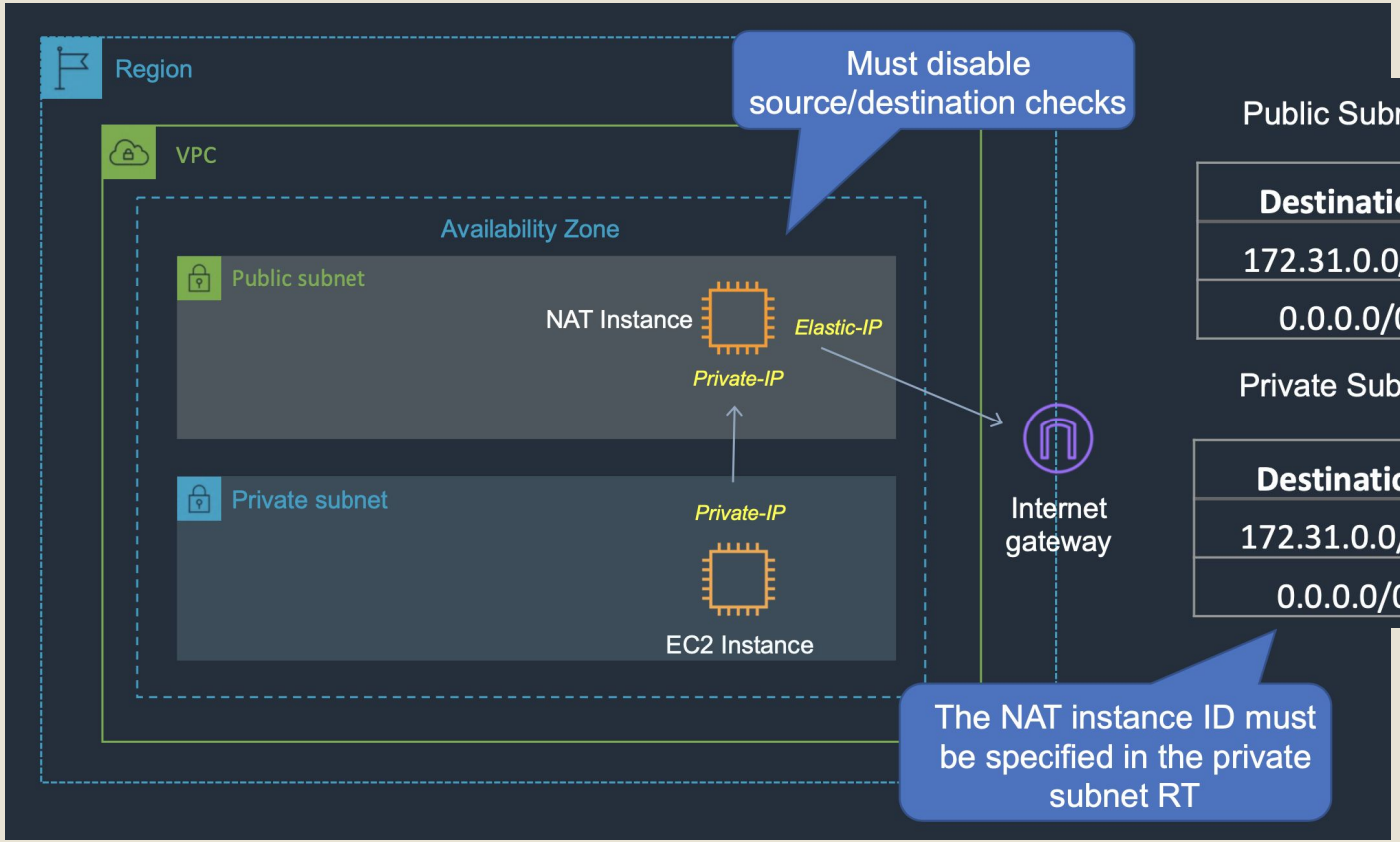
Public Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*igw-id*

Private Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*nat-gateway-id*

Private Subnet with NAT Instance



Public Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*igw-id*

Private Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*nat-instance-id*

Standard Amazon CloudWatch Metrics for EC2

InstanceId	Metric Name
i-0564f898a32460a7c	StatusCheckFailed_System
i-0564f898a32460a7c	StatusCheckFailed_Instance
i-0564f898a32460a7c	StatusCheckFailed
i-0564f898a32460a7c	MetadataNoToken
i-0564f898a32460a7c	NetworkPacketsIn
i-0564f898a32460a7c	NetworkPacketsOut
i-0564f898a32460a7c	CPUUtilization
i-0564f898a32460a7c	NetworkIn
i-0564f898a32460a7c	NetworkOut
i-0564f898a32460a7c	DiskReadBytes
i-0564f898a32460a7c	DiskWriteBytes
i-0564f898a32460a7c	DiskReadOps
i-0564f898a32460a7c	DiskWriteOps
i-0564f898a32460a7c	CPUCreditUsage
i-0564f898a32460a7c	CPUCreditBalance
i-0564f898a32460a7c	CPUSurplusCreditBalance
i-0564f898a32460a7c	CPUSurplusCreditsCharged

There are NO metrics for memory or disk utilization

Custom Amazon CloudWatch Metrics for EC2

- Can publish metrics using the API or AWS CLI
- Example CLI command:

aws cloudwatch **put-metric-data** --metric-name TEST --namespace MyNameSpace --unit Bytes --value 231434333 --dimensions InstanceId=1-23456789,InstanceType=m1.small

- Or you can use the Unified Amazon CloudWatch Agent
- Collects system-level metrics from EC2 and on-premises servers

InstanceId	InstanceType	objectname	Metric Name
i-0564f898a32460a7c	t2.micro	Memory	Memory % Committed Bytes In Use

InstanceId	InstanceType	instance	objectname	Metric Name
i-0564f898a32460a7c	t2.micro	C:	LogicalDisk	LogicalDisk % Free Space

Custom Amazon CloudWatch Metrics for EC2

The unified CloudWatch agent enables you to do the following:

- Collect more system-level metrics from Amazon EC2 instances across operating systems.
  The metrics can include in-guest metrics, in addition to the metrics for EC2 instances.

- Collect system-level metrics from on-premises servers. These can include servers in a hybrid environment as well as servers not managed by AWS

- Retrieve custom metrics from your applications or services using the StatsD and collectd protocols.

- Collect logs from Amazon EC2 instances and on-premises servers, running either Linux or Windows Server

- You can download and install the CloudWatch agent manually using the command line, or you can integrate it with SSM

IAM Policy Example – Allow Full EC2 access in the us-east-2 Region

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": "ec2:*",
            "Resource": "*",
            "Effect": "Allow",
            "Condition": {
                "StringEquals": {
                    "ec2:Region": "us-east-2"
                }
            }
        }
    ]
}
```

IAM Policy Example – Limit Terminating EC2 Instances to an IP Address Range

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["ec2:TerminateInstances"],
            "Resource": ["*"]
        },
        {
            "Effect": "Deny",
            "Action": ["ec2:TerminateInstances"],
            "Condition": {
                "NotIpAddress": {
                    "aws:SourceIp": [
                        "192.0.2.0/24",
                        "203.0.113.0/24"
                    ]
                }
            },
            "Resource": ["*"]
        }
    ]
}
```

AWS EC2 Summary

What is AWS EC2?

- EC2 stands for Elastic Compute Cloud.
- Amazon EC2 is the virtual machine in the Cloud Environment.
- Amazon EC2 provides scalable capacity. Instances can scale up and down automatically based on the traffic.
- You do not have to invest in the hardware.
- You can launch as many servers as you want and you will have complete control over the servers and can manage security, networking, and storage.

Instance Type

- Instance type is providing a range of instance types for various use cases.
- The instance is the processor and memory of your EC2 instance.

EBS Volume:

- EBS Stands for Elastic Block Storage.
- It is the block-level storage that is assigned to your single EC2 Instance.
- It persists independently from running EC2.

o Types of EBS Storage

  - General Purpose (SSD)
  - Provisioned IOPS (SSD)
  - Throughput Optimized Hard Disk Drive
  - Cold Hard Disk Drive
  - Magnetic

AMI

AMI Stands for Amazon Machine Image.

- AMI decides the OS, installs dependencies, libraries, data of your EC2 instances.
- Multiple instances with the same configuration can be launched using a single AMI.

Instance Store

Instance store is the ephemeral block-level storage for the EC2 instance.

- Instance stores can be used for faster processing and temporary storage of the application.

Security Group:

A Security group acts as a virtual firewall for your EC2 Instances.

- It decides the type of port and kind of traffic to allow.
- Security groups are active at the instance level whereas Network ACLs are active at the subnet level.
- Security Groups can only allow but can’t deny the rules.
- The Security group is considered stateful.
- By default, in the outbound rule all traffic is allowed and needs to define the inbound rules.

Tags

Tag is a key-value name you assign to your AWS Resources.

- Tags are the identifier of the resource.
- Resources can be organized well using the tags.

Key Pair

A key pair, consisting of a private key and a public key, is a set of security credentials that you can use to prove your identity while connecting to an instance.

● Amazon EC2 instances use two keys, one is the public key which is attached to your EC2 instance.

● Another is the private key which is with you. You can get access to the EC2 instance only if these keys get matched.

● Keep the private key in a secure place.

Pricing

- You will get different pricing options such as On-Demand, Savings Plan, Reserved Instances, and Spot Instances.

EBS

What's an EBS Volume?

- An EBS (Elastic Block Store) Volume is a network drive you can attach to your instances while they run
- It allows your instances to persist data, even after their termination
- They can only be mounted to one instance at a time (at the CCP level)
- They are bound to a specific availability zone
- Analogy: Think of them as a "network USB stick"
- Free tier: 30 GB of free EBS storage of type General Purpose (SSD) or Magnetic per month

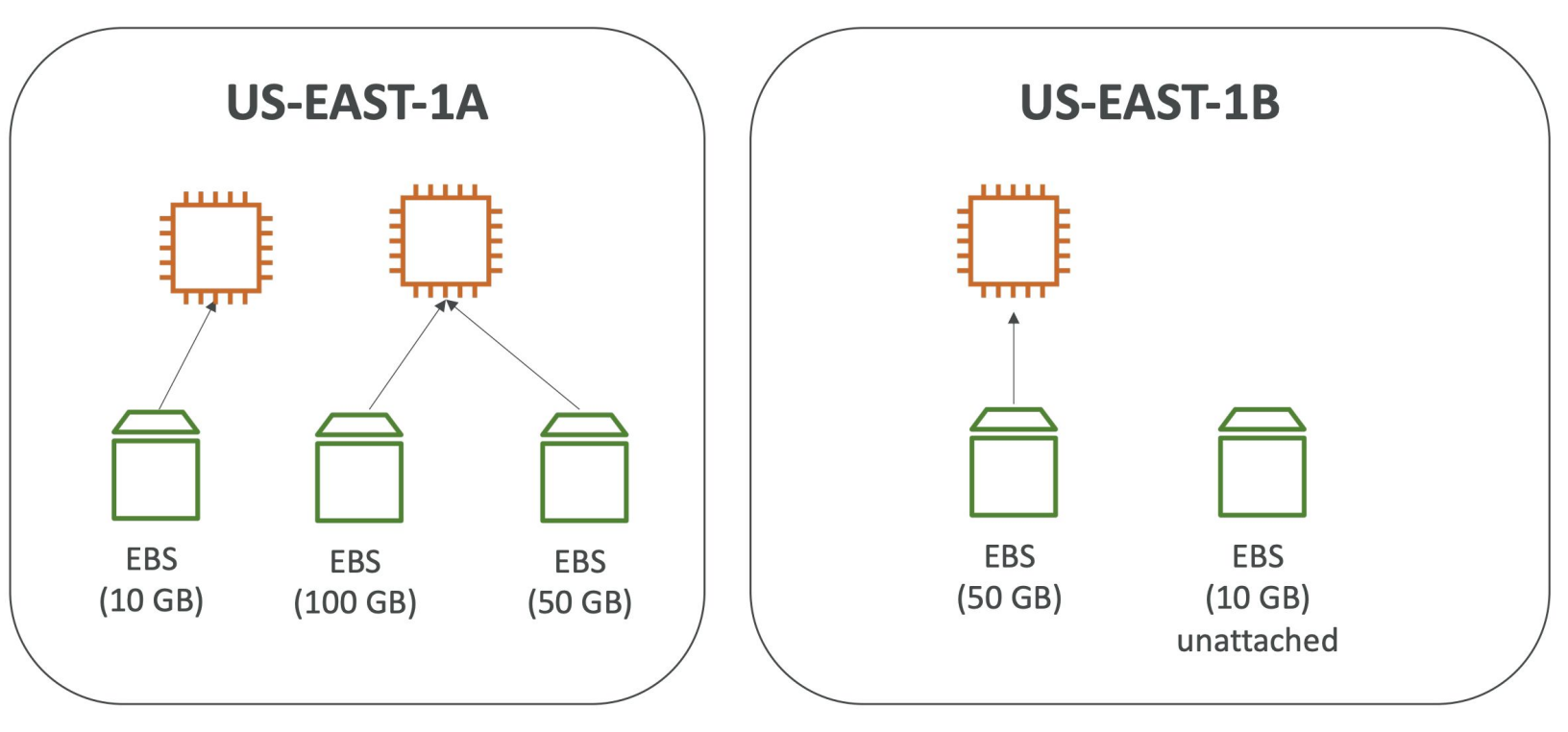
EBS Volume

- ✔ **It's a network drive (i.e. not a physical drive)**
  - It uses the network to communicate the instance, which means there might be a bit of latency
  - It can be detached from an EC2 instance and attached to another one quickly

- ✔ **It's locked to an Availability Zone (AZ)**
  - An EBS Volume in us-east-1a cannot be attached to us-east-1b
  - To move a volume across, you first need to snapshot it

- ✔ **Have a provisioned capacity (size in GBs, and IOPS)**
  - You get billed for all the provisioned capacity
  - You can increase the capacity of the drive over time

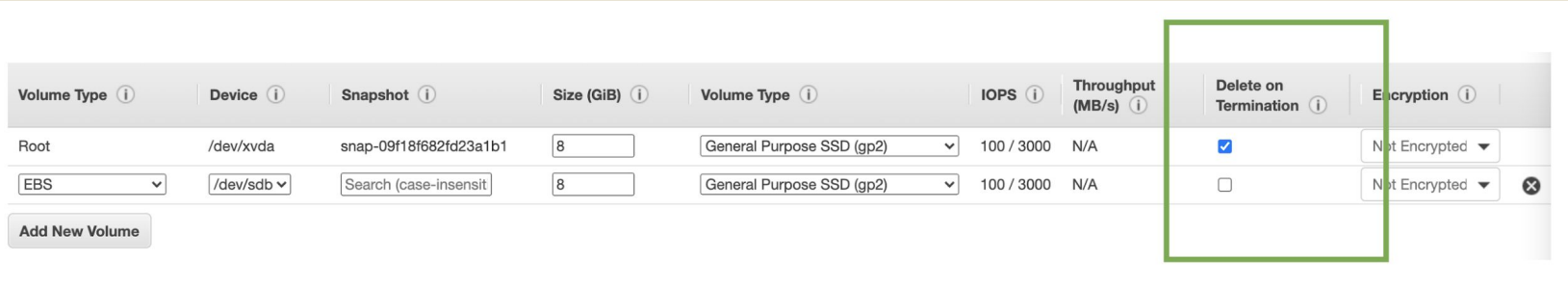
EBS Volume - Example



EBS – Delete on Termination attribute

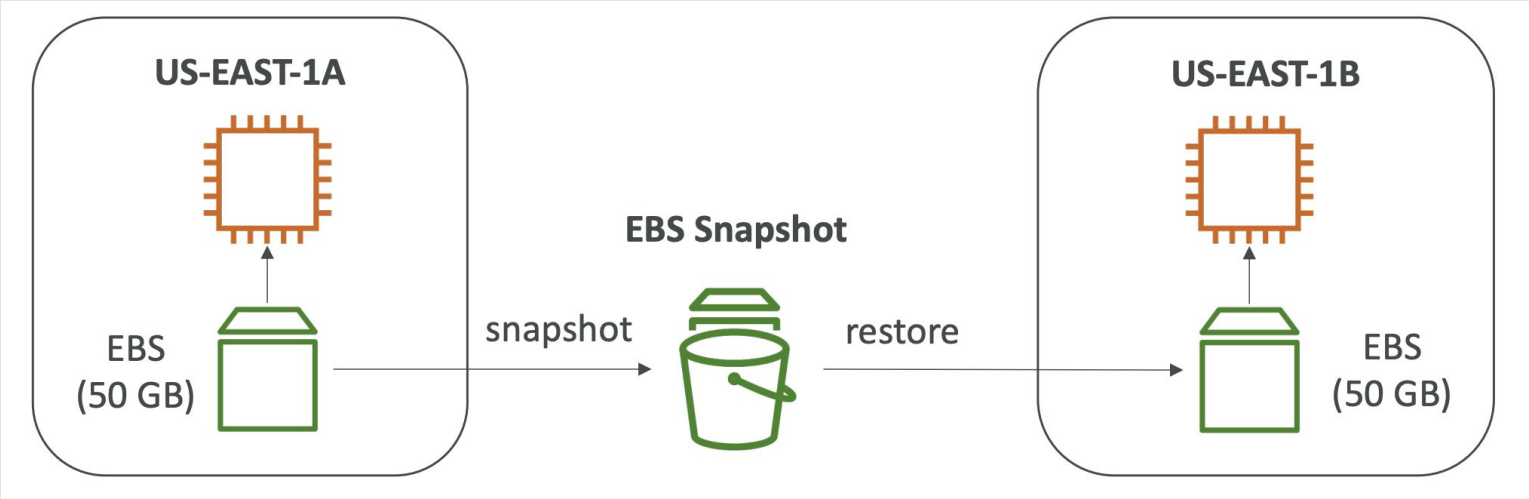
- Controls the EBS behavior when an EC2 instance terminates
  - By default, the root EBS volume is deleted (attribute enabled)
  - By default, any other attached EBS volume is not deleted (attribute disabled)

- This can be controlled by the AWS console / AWS CLI
- Use case: preserve root volume when instance is terminated



EBS Snapshots

- Make a backup (snapshot) of your EBS volume at a point in time
- Not necessary to detach volume to do snapshot, but recommended
- Can copy snapshots across AZ or Region



EBS Snapshots Features

- **EBS Snapshot Archive**

- Move a Snapshot to an ”archive tier” that is 75% cheaper
- Takes within 24 to 72 hours for restoring the archive

- **Recycle Bin for EBS Snapshots**

- Setup rules to retain deleted snapshots so you can recover them after an accidental deletion
- Specify retention (from 1 day to 1 year)

- **Fast Snapshot Restore (FSR)**

- Force full initialization of snapshot to have no latency on the first use ($$$)

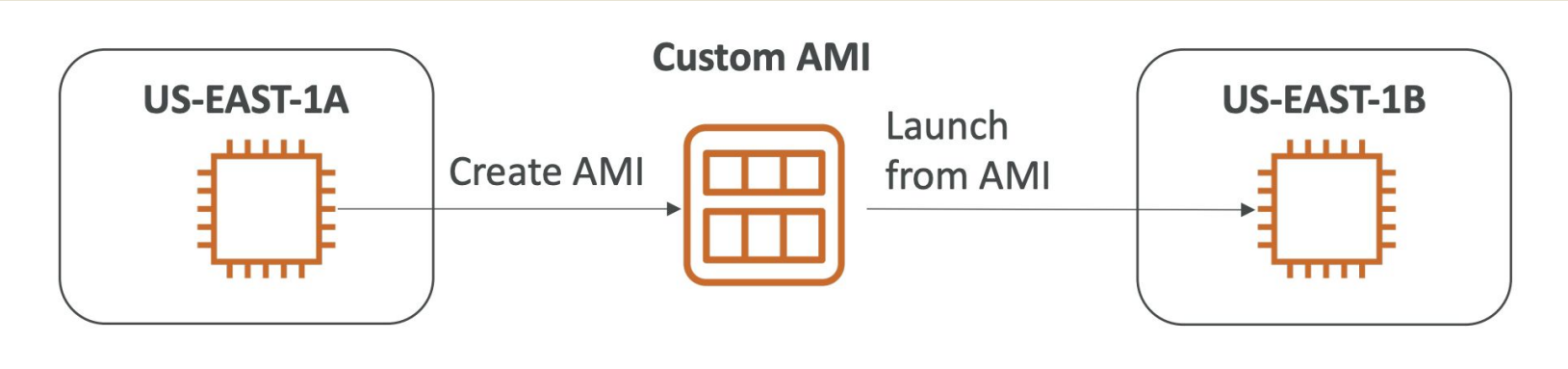
AMI Overview

- AMI = Amazon Machine Image
- AMI are a customization of an EC2 instance
  - You add your own software, configuration, operating system, monitoring…
  - Faster boot / configuration time because all your software is pre-packaged

- AMI are built for a specific region (and can be copied across regions)
- You can launch EC2 instances from:
  - A Public AMI: AWS provided
  - Your own AMI: you make and maintain them yourself
  - An AWS Marketplace AMI: an AMI someone else made (and potentially sells)

AMI Process (from an EC2 instance)

- Start an EC2 instance and customize it
- Stop the instance (for data integrity)
- Build an AMI – this will also create EBS snapshots
- Launch instances from other AMIs



EC2 Instance Store

- EBS volumes are network drives with good but “limited” performance
- If you need a high-performance hardware disk, use EC2 Instance Store
- Better I/O performance
- EC2 Instance Store lose their storage if they’re stopped (ephemeral)
- Good for buffer / cache / scratch data / temporary content
- Risk of data loss if hardware fails
- Backups and Replication are your responsibility

EBS Volume Types

- EBS Volumes come in 6 types
- gp2 / gp3 (SSD): General purpose SSD volume that balances price and performance for a wide variety of workloads
- io1 / io2 (SSD): Highest-performance SSD volume for mission-critical low-latency or high-throughput workloads
- st1 (HDD): Low cost HDD volume designed for frequently accessed, throughput- intensive workloads
- sc1 (HDD): Lowest cost HDD volume designed for less frequently accessed workloads

- EBS Volumes are characterized in Size | Throughput | IOPS (I/O Ops Per Sec)
- When in doubt always consult the AWS documentation – it's good!
- Only gp2/gp3 and io1/io2 can be used as boot volumes

EBS Volume Types Use cases

General Purpose SSD

- Cost effective storage, low-latency
- System boot volumes, Virtual desktops, Development and test environments
- 1 GiB - 16 TiB
- gp3:

  - Baseline of 3,000 IOPS and throughput of 125 MiB/s

  - Can increase IOPS up to 16,000 and throughput up to 1000 MiB/s independently

- gp2:

  - Small gp2 volumes can burst IOPS to 3,000

  - Size of the volume and IOPS are linked, max IOPS is 16,000

  - 3 IOPS per GB, means at 5,334 GB we are at the max IOPS

EBS Volume Types Use cases

Provisioned IOPS (PIOPS) SSD

- Critical business applications with sustained IOPS performance
- Or applications that need more than 16,000 IOPS
- Great for databases workloads (sensitive to storage perf and consistency)
- io1/io2 (4 GiB - 16 TiB):

- Max PIOPS: 64,000 for Nitro EC2 instances & 32,000 for other
- Can increase PIOPS independently from storage size
- io2 have more durability and more IOPS per GiB (at the same price as io1)

- io2 Block Express (4 GiB – 64 TiB):

- Sub-millisecond latency
- Max PIOPS: 256,000 with an IOPS:GiB ratio of 1,000:1

- Supports EBS Multi-attach

EBS Volume Types Use cases

Hard Disk Drives (HDD)

- Cannot be a boot volume
- 125 GiB to 16 TiB
- Throughput Optimized HDD (st1)
  - Big Data, Data Warehouses, Log Processing
  - Max throughput 500 MiB/s – max IOPS 500
- Cold HDD (sc1):
  - For data that is infrequently accessed
  - Scenarios where lowest cost is important
  - Max throughput 250 MiB/s – max IOPS 250

EBS – Volume Types Summary

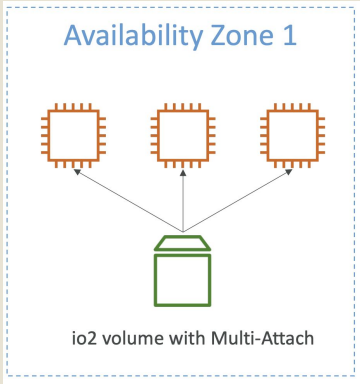
	General Purpose SSD		Provisioned IOPS SSD		
Volume type	gp3	gp2	io2 Block Express ‡	io2	io1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.999% durability (0.001% annual failure rate)		99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	• Low-latency interactive apps • Development and test environments		Workloads that require sub-millisecond latency, and sustained IOPS performance or more than 64,000 IOPS or 1,000 MiB/s of throughput	• Workloads that require sustained IOPS performance or more than 16,000 IOPS • I/O-intensive database workloads	
Volume size	1 GiB - 16 TiB		4 GiB - 64 TiB	4 GiB - 16 TiB	
Max IOPS per volume (16 KiB I/O)	16,000		256,000	64,000 †	

	Throughput Optimized HDD	Cold HDD
Volume type	st1	sc1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	• Big data • Data warehouses • Log processing	• Throughput-oriented storage for data that is infrequently accessed • Scenarios where the lowest storage cost is important
Volume size	125 GiB - 16 TiB	125 GiB - 16 TiB
Max IOPS per volume (1 MiB I/O)	500	250
Max throughput per volume	500 MiB/s	250 MiB/s
Amazon EBS Multi-attach	Not supported	Not supported
Boot volume	Not supported	Not supported

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html

EBS Multi-Attach – io1/io2 family

- Attach the same EBS volume to multiple EC2 instances in the same AZ
- Each instance has full read & write permissions to the high-performance volume
- Use case:
- Achieve higher application availability in clustered Linux applications (ex: Teradata)
- Applications must manage concurrent write operations
- Up to 16 EC2 Instances at a time
- Must use a file system that’s cluster-aware (not XFS, EX4, etc…)



THANK YOU!