

AWS Certified Solutions Architect Associate

By Stéphane Maarek



COURSE →

[screencast-a-associate-coupon](#)



EXTRA PRACTICE EXAMS →

Disclaimer: These slides are copyrighted and strictly for personal use only

- This document is reserved for people enrolled into the [Ultimate AWS Solutions Architect Associate Course](#)
- Please do not share this document, it is intended for personal use and exam preparation only, thank you.
- If you've obtained these slides for free on a website that is not the course's website, please reach out to piracy@datacumulus.com. Thanks!
- Best of luck for the exam and happy learning!

Table of Contents

- [Getting Started with AWS](#)
- [AWS Identity & Access Management \(AWS IAM\)](#)
- [Amazon EC2 - Basics](#)
- [Amazon EC2 - Associate](#)
- [Amazon EC2 - Instance Storage](#)
- [High Availability & Scalability](#)
- [RDS, Aurora & ElastiCache](#)
- [Amazon Route 53](#)
- [Classic Solutions Architecture](#)
- [Amazon S3](#)

Table of Contents

- [Amazon S3 - Advanced](#)
- [Amazon S3 - Security](#)
- [CloudFront & Global Accelerator](#)
- [AWS Storage Extras](#)
- [AWS Integration & Messaging](#)
- [Containers on AWS](#)
- [Serverless Overview](#)
- [Serverless Architectures](#)
- [Databases in AWS](#)
- [Data & Analytics](#)

Table of Contents

- [Machine Learning](#)
- [AWS Monitoring, Audit & Performance](#)
- [Advanced Identity in AWS](#)
- [AWS Security & Encryption](#)
- [Amazon VPC](#)
- [Disaster Recovery & Migrations](#)
- [More Solutions Architecture](#)
- [Other Services](#)
- [White Papers & Architectures](#)
- [Exam Preparation](#)
- [Congratulations](#)

AWS Certified Solutions Architect Associate Course SAA-C03

Welcome! We're starting in 5 minutes



- We're going to prepare for the Solutions Architect exam - SAA-C03
- It's a challenging certification, so this course will be long and interesting
- Basic IT knowledge is necessary
- This course contains videos...
 - From the Cloud Practitioner, Developer and SysOps course - shared knowledge
 - Specific to the Solutions Architect exam - exciting ones on architecture!
- We will cover over 30 AWS services
- AWS / IT Beginners welcome! (but take your time, it's not a race)

My SAA-C03 certification: 96.1%

AWS Certified Solutions Architect - Associate

Notice of Exam Results

Candidate: Stephane MAAREK	Exam Date: Sep 02, 2022
Candidate ID: AWS [REDACTED]	Registration Number: [REDACTED]
Candidate Score: 961	Pass/Fail: PASS

About me

- I'm Stephane!
- Worked as in IT consultant and AWS Solutions Architect, Developer & SysOps
- Worked with AWS many years: built websites, apps, streaming platforms
- Veteran Instructor on AWS (Certifications, CloudFormation, Lambda, EC2...)
- You can find me on
 - GitHub: <https://github.com/simplesteph>
 - LinkedIn: <https://www.linkedin.com/in/stephanemaarek>
 - Medium: <https://medium.com/@stephane.maarek>
 - Twitter: <https://twitter.com/stephanemaarek>



4.7 Instructor Rating
 473,642 Reviews
 1,553,489 Students
 39 Courses

What's AWS?



- AWS (Amazon Web Services) is a Cloud Provider
- They provide you with servers and services that you can use on demand and scale easily
- AWS has revolutionized IT over time
- AWS powers some of the biggest websites in the world
 - Amazon.com
 - Netflix

What we'll learn in this course (and more!)

Amazon
EC2

Amazon ECR



Amazon ECS

AWS Elastic
BeanstalkAWS
Lambda

Auto Scaling



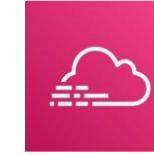
IAM



AWS KMS

Amazon
S3Amazon
SESAmazon
RDSAmazon
AuroraAmazon
DynamoDBAmazon
ElastiCacheAmazon
SQSAmazon
SNS

AWS Step Functions

Amazon
CloudWatchAWS
CloudFormationAWS
CloudTrailAmazon API
GatewayElastic Load
BalancingAmazon
CloudFrontAmazon
KinesisAmazon
Route 53

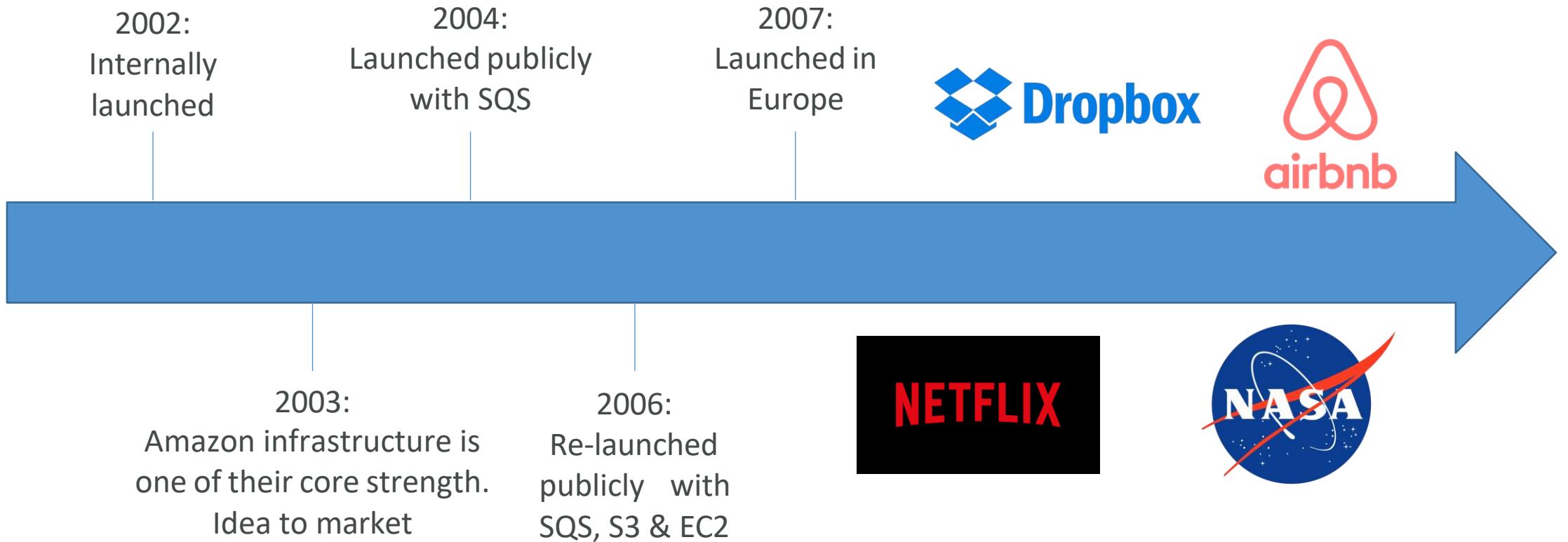
Navigating the AWS spaghetti bowl



Udemy Tips

Getting started with AWS

AWS Cloud History



AWS Cloud Number Facts

- In 2023, AWS had \$90 billion in annual revenue
- AWS accounts for 31% of the market in Q1 2024 (Microsoft is 2nd with 25%)
- Pioneer and Leader of the AWS Cloud Market for the 13th consecutive year
- Over 1,000,000 active users

Figure 1: Magic Quadrant for Strategic Cloud Platform Services



Gartner Magic Quadrant

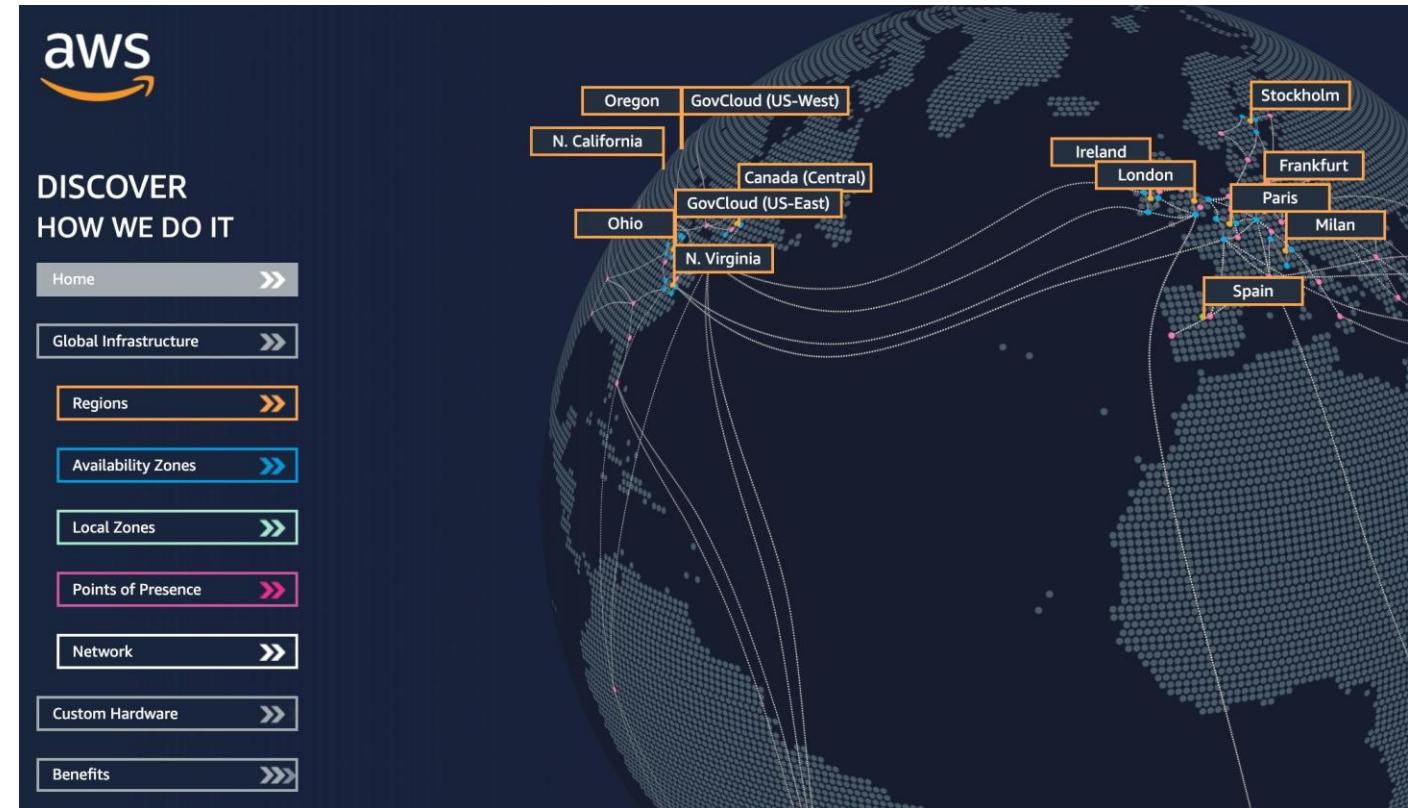
AWS Cloud Use Cases

- AWS enables you to build sophisticated, scalable applications
- Applicable to a diverse set of industries
- Use cases include
 - Enterprise IT, Backup & Storage, Big Data analytics
 - Website hosting, Mobile & Social Apps
 - Gaming



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/>

US East (N. Virginia) us-east-1

US East (Ohio) us-east-2

US West (N. California) us-west-1

US West (Oregon) us-west-2

Africa (Cape Town) af-south-1

Asia Pacific (Hong Kong) ap-east-1

Asia Pacific (Mumbai) ap-south-1

Asia Pacific (Seoul) ap-northeast-2

Asia Pacific (Singapore) ap-southeast-1

Asia Pacific (Sydney) ap-southeast-2

Asia Pacific (Tokyo) ap-northeast-1

Canada (Central) ca-central-1

Europe (Frankfurt) eu-central-1

Europe (Ireland) eu-west-1

Europe (London) eu-west-2

Europe (Paris) eu-west-3

Europe (Stockholm) eu-north-1

Middle East (Bahrain) me-south-1

South America (São Paulo) sa-east-1

How to choose an AWS Region?

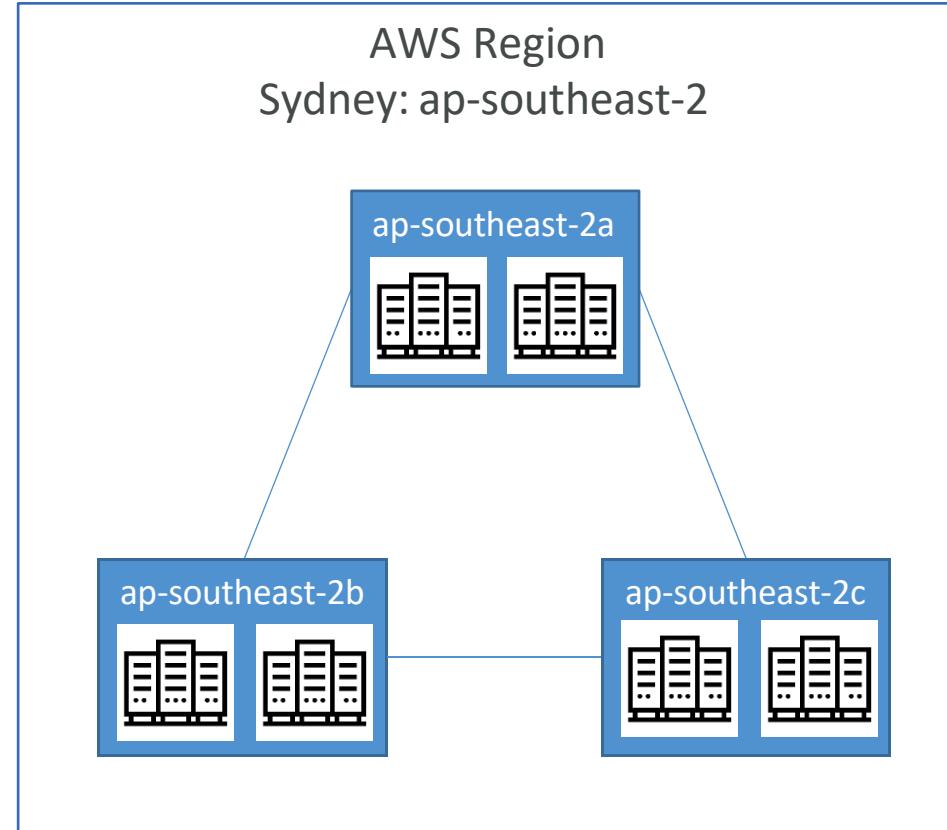
If you need to launch a new application,
where should you do it?



- **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

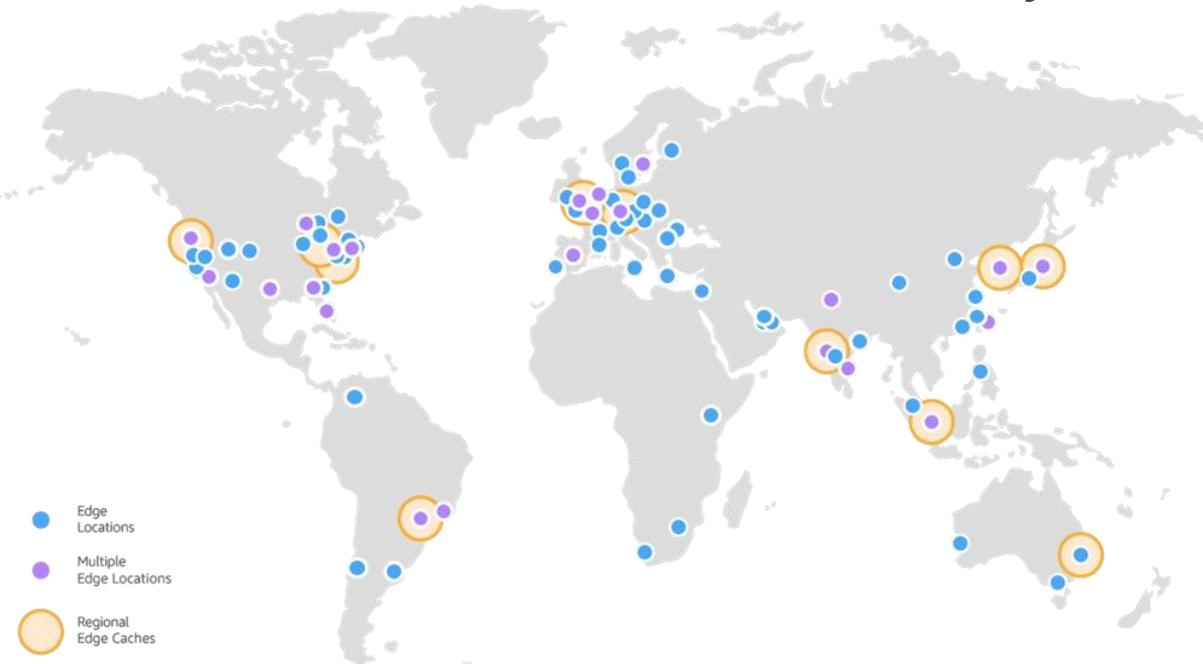
AWS Availability Zones

- Each region has many availability zones (usually 3, min is 3, 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)

- Amazon has 400+ Points of Presence (400+ Edge Locations & 10+ Regional Caches) in 90+ cities across 40+ countries
- Content is delivered to end users with lower latency



<https://aws.amazon.com/cloudfront/features/>

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>



AWS Identity and Access Management (AWS IAM)

IAM: Users & Groups



- IAM = Identity and Access Management, Global service
- Root account created by default, shouldn't be used or shared
- Users are people within your organization, and can be grouped
- Groups only contain users, not other groups
- Users don't have to belong to a group, and user can belong to multiple groups



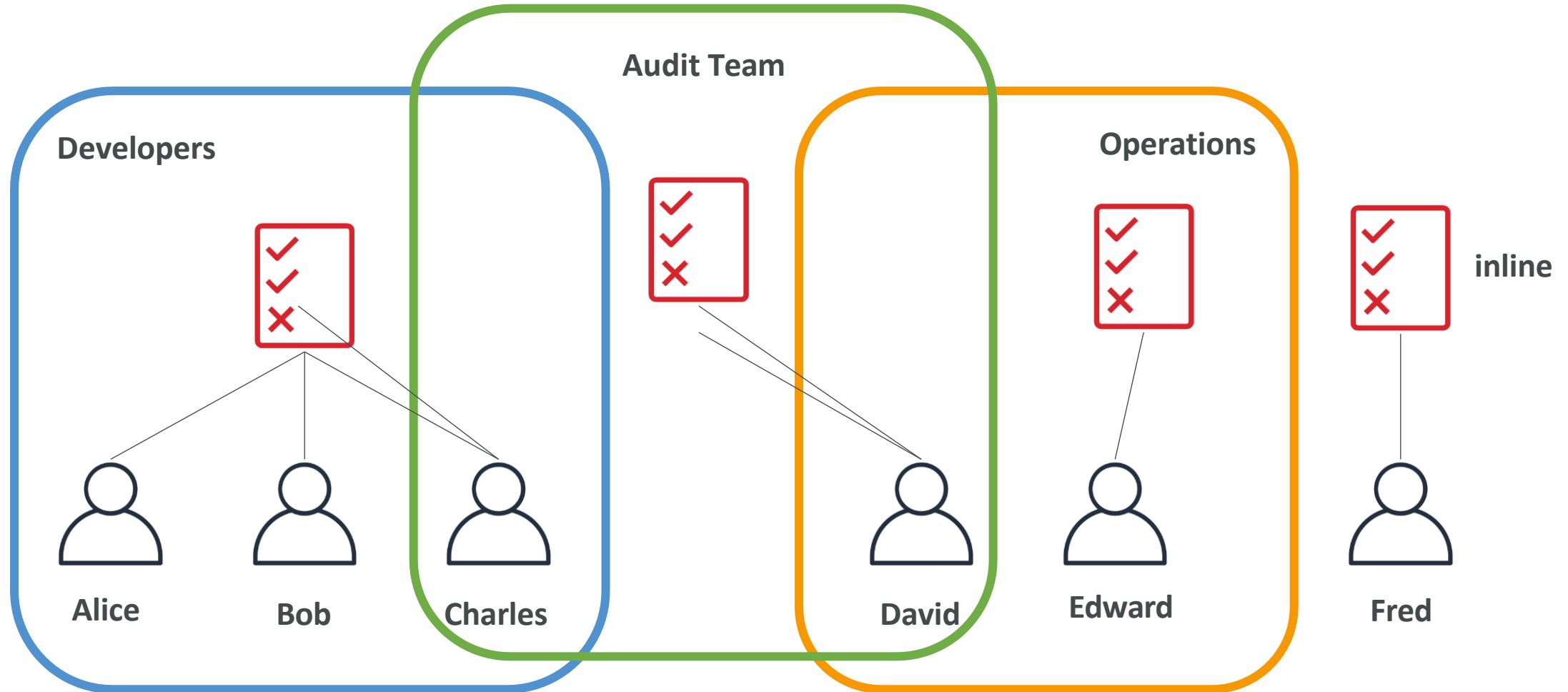
IAM: Permissions

- Users or Groups can be assigned JSON documents called policies
- These policies define the permissions of the users
- In AWS you apply the least privilege principle: don't give more permissions than a user needs

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": "ec2:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": "elasticloadbalancing:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": [  
                "cloudwatch>ListMetrics",  
                "cloudwatch:GetMetricStatistics",  
                "cloudwatch:Describe*"  
            ],  
            "Resource": "*"  
        }  
    ]  
}
```



IAM Policies inheritance



IAM Policies Structure

- Consists of
 - Version: policy language version, always include “2012-10-17”
 - Id: an identifier for the policy (optional)
 - Statement: one or more individual statements (required)
- Statements consists of
 - Sid: an identifier for the statement (optional)
 - Effect: whether the statement allows or denies access (Allow, Deny)
 - Principal: account/user/role to which this policy applied to
 - Action: list of actions this policy allows or denies
 - Resource: list of resources to which the actions applied to
 - Condition: conditions for when this policy is in effect (optional)

```
{  
  "Version": "2012-10-17",  
  "Id": "S3-Account-Permissions",  
  "Statement": [  
    {  
      "Sid": "1",  
      "Effect": "Allow",  
      "Principal": {  
        "AWS": ["arn:aws:iam::123456789012:root"]  
      },  
      "Action": [  
        "s3:GetObject",  
        "s3:PutObject"  
      ],  
      "Resource": ["arn:aws:s3:::mybucket/*"]  
    }  
  ]  
}
```

IAM - Password Policy

- Strong passwords = higher security for your account
- In AWS, you can setup a password policy:
 - Set a minimum password length
 - Require specific character types:
 - including uppercase letters
 - lowercase letters
 - numbers
 - non-alphanumeric characters
 - Allow all IAM users to change their own passwords
 - Require users to change their password after some time (password expiration)
 - Prevent password re-use

Multi Factor Authentication - MFA



- Users have access to your account and can possibly change configurations or delete resources in your AWS account
- You want to protect your Root Accounts and IAM users
- MFA = password you know + security device you own



Alice

Password

+



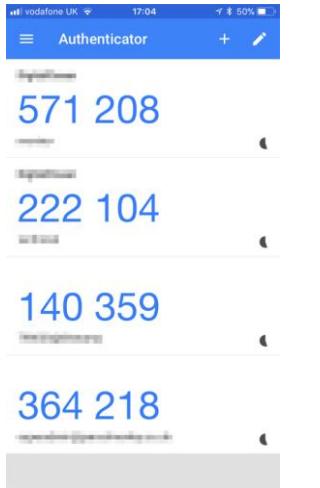
=>

Successful login

- Main benefit of MFA:
if a password is stolen or hacked, the account is not compromised

MFA devices options in AWS

Virtual MFA device



Google Authenticator
(phone only)

Support for multiple tokens on a single device.



Authy
(phone only)

Universal 2nd Factor (U2F) Security Key



YubiKey by Yubico (3rd party)

Support for multiple root and IAM users
using a single security key

MFA devices options in AWS

Hardware Key Fob MFA Device



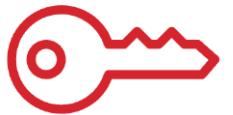
Provided by Gemalto (3rd party)

Hardware Key Fob MFA Device for AWS GovCloud (US)



Provided by SurePassID (3rd party)

How can users access AWS ?



- To access AWS, you have three options:
 - AWS Management Console (protected by password + MFA)
 - AWS Command Line Interface (CLI): protected by access keys
 - AWS Software Developer Kit (SDK) - for code: protected by access keys
- Access Keys are generated through the AWS Console
- Users manage their own access keys
- Access Keys are secret, just like a password. Don't share them
- Access Key ID ~ = username
- Secret Access Key ~ = password

Example (Fake) Access Keys

Access keys

Use access keys to make secure REST or HTTP Query protocol requests to AWS service APIs. For your protection, you should never share your secret keys with anyone. As a best practice, we recommend frequent key rotation. [Learn more](#)

[Create access key](#)

Access key ID	Created	Last used	Status	
AKIASK4E37PV4TU3RD6C	2020-05-25 15:13 UTC+0100	N/A	Active	Make inactive X

- Access key ID: AKIASK4E37PV4983d6C
- Secret Access Key: AZPN3z0jWozWCndljhB0Unh8239a1bzBzO5fqkZq
- Remember: don't share your access keys

What's the AWS CLI?

- A tool that enables you to interact with AWS services using commands in your command-line shell
- Direct access to the public APIs of AWS services
- You can develop scripts to manage your resources
- It's open-source <https://github.com/aws/aws-cli>
- Alternative to using AWS Management Console

```
→ ~ aws s3 cp myfile.txt s3://ccp-mybucket/myfile.txt
upload: ./myfile.txt to s3://ccp-mybucket/myfile.txt
→ ~ aws s3 ls s3://ccp-mybucket
2021-05-14 03:22:52          0 myfile.txt
→ ~ █
```

What's the AWS SDK?



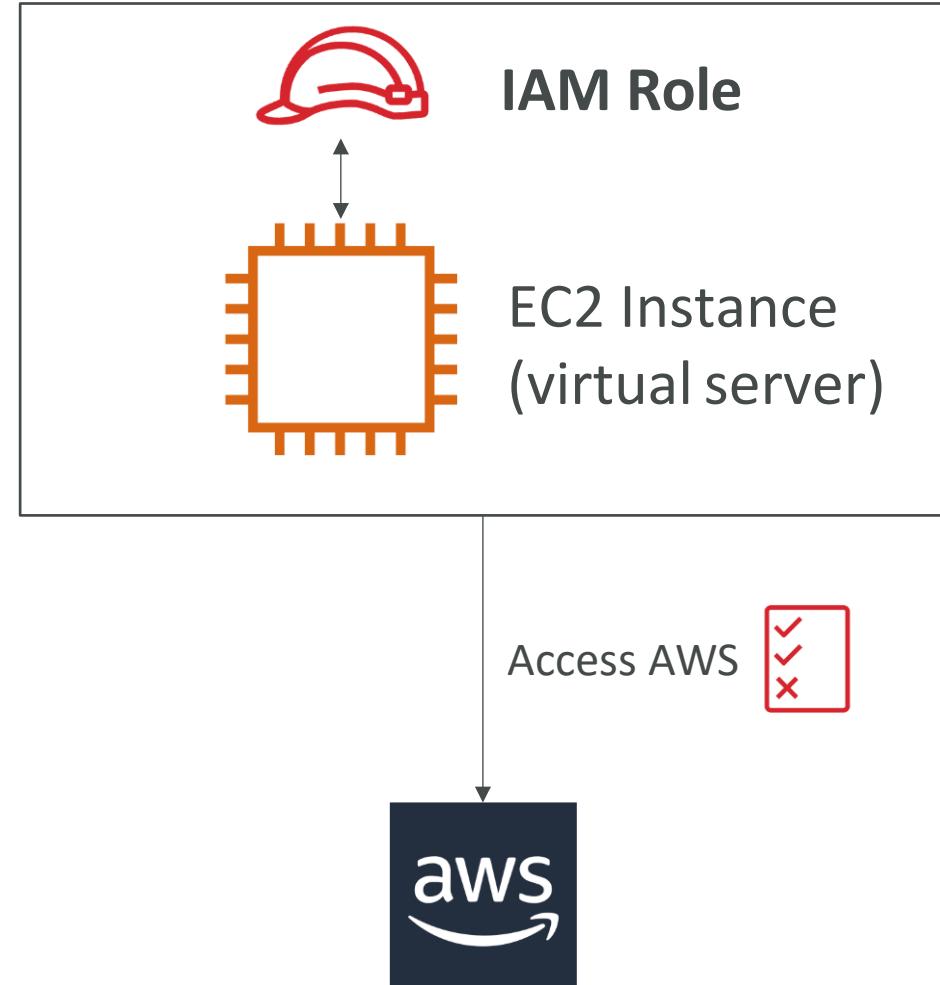
- AWS Software Development Kit (AWS SDK)
- Language-specific APIs (set of libraries)
- Enables you to access and manage AWS services programmatically
- Embedded within your application
- Supports
 - SDKs (JavaScript, Python, PHP, .NET, Ruby, Java, Go, Node.js, C++)
 - Mobile SDKs (Android, iOS, ...)
 - IoT Device SDKs (Embedded C, Arduino, ...)
- Example: AWS CLI is built on AWS SDK for Python



Your Application

IAM Roles for Services

- Some AWS service will need to perform actions on your behalf
- To do so, we will assign permissions to AWS services with IAM Roles
- Common roles:
 - EC2 Instance Roles
 - Lambda Function Roles
 - Roles for CloudFormation



IAM Security Tools

- IAM Credentials Report (account-level)
 - a report that lists all your account's users and the status of their various credentials
- IAM Access Advisor (user-level)
 - Access advisor shows the service permissions granted to a user and when those services were last accessed.
 - You can use this information to revise your policies.

IAM Guidelines & Best Practices



- Don't use the root account except for AWS account setup
- One physical user = One AWS user
- Assign users to groups and assign permissions to groups
- Create a strong password policy
- Use and enforce the use of Multi Factor Authentication (MFA)
- Create and use Roles for giving permissions to AWS services
- Use Access Keys for Programmatic Access (CLI / SDK)
- Audit permissions of your account using IAM Credentials Report & IAM Access Advisor
- Never share IAM users & Access Keys

So Hiiiii, yesterday's messages were 1000% true and all those detailed clarifications were sent to prevent you from getting a heart attack reg Hope and me getting Hurt

Extremely sorry if I caused any sort of heart heaviness or pressure or anything at all..Any inconvenience to your daily life if i caused naa, i sincerely apologise and it was never ever my intention!.. Extremely sorry!

if all of this feels way too overwhelming to you and too much to process, I'll simplify it a bit to you

Your closest friend fell for you after knowing your complete personality - he started to like you much more and admired every single thing you did sweetly ..that's it

Idhula edhvume bp raise aaga illa and with time and perspective, you will realise it's just lighthearted..
and just nee enna decide panaalum in the end , oru stranger ah vida , he ll understand it soo well!

Just relaxxx, you don't have to reply anything or decide anything..just listen to what I am saying and let's carry on with our studies

After you have read this, just let me know that you have fully and patiently read every single word..that's all ..padishten nu mattum reply pannu podhum lol

Clarification clarification clarification nu solli main point ah poda miss pannitten 

Note : these are not compliments..these are facts ma!
I request you read every single word in a light hearted and sweet way without bp increasing..
ahn vaazhthukkal vaazhthukkal..
aarambikkaalaama?

That kanmani anbodu song la "Unnai yenni paarkayil kavithai kottudhuu, adhai yezhudha nenaikayil vaarthai muttudhu" is literally how I am feeling right now while writing this

Seee first unna paakum bhodhey I did have little sparks flying out of me because I was not expecting that look from a Melmaruvathur ponnu haha..

Honestly, From the beginning, I had some special place for you..But right from the first day of college, I was obsessed with youtube..so I didn't even think about anything else

But I swear appolerundhey I won't attend anyone's calls or messages..kanish laam appo dhaan was starting to be friends and even he was pissed off..

But I used to call back and message only you back whenever I saw your message

Some subconscious additional care and affection I had - andha time la adhukku enna paer vekkardhu nu theriyaama I called it "sister" and there is not a single day that goes by that i don't regret calling you that when I didn't even know about you

Actually naan oru vishayam solla maranten..actually nee nadula relationship time la enkitta pesaama irundha la..

Appo sometimes, I used to listen to sad songs anolll 😭 and feel pannitutu iruppen nee yen enkitta pesardhu stopped nu .. as a friend..

Cause you were the first friend I made and you were special to me..

I even remember one of the songs i listened to..It's called "nanbiye" from Teddy nu oru arya movie

So the foundation was there from the very beginning, this wasn't something that suddenly jumped out of nowhere..

My Sparks for you restarted for you exactly on your birthday when we had a little time to talk outside Dominos.. Something about the way you spoke(like romba calm ah) I started to like that day..i liked it as a friend

Then I started to notice other positive qualities about you - like that audition pona modhu, how you told madhushree - "apdi laam pesaadha and expression kudukaadha madhushree avangalukku hurt aagum"..
Adhuvum you told her very softly and sweetly - sonna vidham ku hats off

I feel like I am actually getting to know a person called Sakthi priya only from December 2024.. Adhuku munaadi all surface level and education things dhaan i knew about you...

Andha time la I used to discuss with kanish about one of your one positive qualities..

"Indha Harini ku oru vishayam or oru technology or skill Theriyala naa kooda "she is willing to learn and do it"..that was a very rare quality we did not see in other girls here..

Enakku kathukudunga or guide pannunga or video anuppunga, I ll learn and do nu you used to say..adhellam we really admired a

lot compared to others who are summave wasting time and
saying adhellam ennaku theriyaadhu

See nee enna dhaan time waste pannitaalum first 2 years
la...your dedication and Interest to learn something and correct
your mistake and Inime padikka pore apdindra andha thought I
really respect and admire...

That willingness to learn!

Thappu panravan manushan

Thappu othukuravan periya manushan..

Thappu senjittom Naan Inime epdi maara pore paarunga! and

elaathayum seri panna porenu solravan Maa-Manithan and
that's you!!

Indha eduthula thappu naa periya word laam illa and I am not
indicating that nee thappu pannittaa...andha time waste aana
years ah dhaan solren

I have spoken with many girls..Avanga veetla evloo kashtam
irundhaalum, they ll still be scrolling reels and endha oru
padippula interest ilaama they ll be there and avangala ellam
paarkum bhodhu, the amount of responsible thoughts you have is
astounding...

Don't udane bring back the first 2 years la nee pannadhu ellam..
The amount of responsibility you took after all that and said "en
veetla naan vandhu aaganum career la mela" - adhellam i really
admired... how much family responsibility eduthu nee porupaana
ponnu maari pesura nu...that thought and words alone deserves
claps and whistles 🙌 🎉

So un mela oru nulla opinion was always growing on me

Poga poga I realised that Nee oru maari cuteness overloaded -
times infinity maari elaame panra and irukka haha

how shy you get when someone compliments youu , the way you vekkapattufy and blush is sooo cutee!!

And the way you laughhh!!

Ohh my Goddd please stop stealing my heart!! 😊

There was this one time outside Tools class and I saw you and heard you laughing at edho aakash sonnadhukku, and goddd it was soo enchanting and beautiful that I fell!

#SpellBound!!!

Actually ungalukku sadness suthama set aagala mamm!!..Nee laam eppome sirishteyy iruuu..adhaan azhaga irukku un cute face

ku!

Let's come to the most powerful magical spell you cast - the way you talk.. I think there is something fundamentally wrong with people who said something bad about how much you talk..

I literally feel like your voice and way of talking is soo therapeutic and calm and can heal literally anything...

You should be sent to jail for asking the question : Am I disturbing you or talking soo much?

Cause it's that much sweet! Sometimes in lab and in class, just listening to you talk to someone else also feels so sweet to hear!

I want to listen to you talk all the time , share every single incident you experienced and about every single topic

I specifically like even all your hmm variations
(imagine your voice)

Normal hmm - "hmm"

Sad and problem la maatikitta hmm - "hm.. hm.. hm.. hm"

Nakkalaa solra hmm - aaaaaaaaaan type la vara
"uhmmmuuhmmm"

And end of topic hmm - "hmmmmmmmmmm...."

And in addition to all this, how you pronounce certain words..
Like "enaachiii"..and "naan ippoo enna pannanumm"....
Sooo cuteeeee yaaaaa!! How can someone not fall for this!!
Manushane illa vizhaadhavan laam

Even the way you say "semmmaaaa po" is actually semmaaaa
dhaar...Nee semmaaaa ponnu dhhaar

The sweetest of them all - "avangalukkuuu edhaa badd ah feel aagumaaa"...do you even realise the beauty of this dialogue?..Not only is it cute with your tone, but it reveals something fundamentally beautiful about your soul that you care deeply about others and you don't want to hurt them..Do you even know how rare this is??

Literally everything about you is sweet..
And ohh yesss –

How the hellll am I supposedddd to not fall for someoneee who is
soo cuteee even when they are angryyyyy?!?????!!!!

Let me demonstrate

Hear this in your angry voice - "Heyyy, Naan unna pathi apdi
sollavey illa pothumaaa?" 😡

And then shortly after..

"Seri sorry yen mela dhaan thappu".....even when you are not fully convinced yet



Eyyyyy enga Irundhu maa vandha nee huh?? 🤣🤣🤣🤣
🤣🤣🤣 Solluuuu..

idhukkum vandhu sorry nu kekura aalu nee 🤣🤣👊👊

Epdiyaapattavanum uzhundhuruvaan maa unakku.. Falling for you was the simplest thing!!

Even your kovam is not something that I am able to take offence for, because your voice is all high pitched and you are thinking you are kovam pattufying and people are just like soo sweet laaa...

Even the way you type on a keyboard is cuteeeeee, In the last lab class, I was literally talking about something and completely forgot about it mid sentence when I saw the way you were typing on your keyboard..like adhula kooda only you can bee thiss level of cute..
Like oru maari childlike typing one by one was soooo cuteee

From the way you give slightly disappointed expressions, From the way you walk when you are tired, from the way you flip your hair to the way you turn your head, i have completely fallen...
Aala udunga I surrenderrrr, naan uzhuntennn podhumaa? 😊

You are like that real life version of Hasini in santosh subramaniam..enna difference naa you are 10 times more intelligent and smarter than her..adhaan difference..Not even joking , Nee eppome actual ah ve topper dhaan..its a factt...if you fully focus, there is nothing you cannot learn and achieve

You are alsoo having fun with life with sooo many pranks and

that's really funnn!!

I got to saw how happy and a fun person you were before the whole pipe phase in your life..I saw how social and how well you speak with people konjam konjam i observed and I can only imagine how happy and how different of a person you will be when you are treated in the right way!

You are one of the gentlest and purest souls I have ever met in my life!

I ll tell you why..

Nee eppo nee pannadhu thappu nu honest ah oru vishayathukku
feel panriyooo, nee eppo thappey Pannala naa kooda oruthanga
kitta sorry kekuriyooo, nee eppo mathavanga hurt aaga
koodaadhu and bad ah feel panna koodaadhu nu nenaikuriyooo,
appove it implies your heart is beautiful!

Nee yaarukkum kettadhu nenaikalaa

Dark topics laam please don't consider even as a topic.. adhelam
jokes category la varudhu..already neraiyya explain
panniyaachh..u r becoming more open minded slowly

Every human makes mistakes..but how much you care about

your close ones and how you want to correct your mistakes instead of neglecting and forever staying guilty is what makes you different

How much efforts you put in to change yourself when someone says un behaviour is hurting them speaks miles about you!

I have seen sooo many incidents!.

Communicate pannum bhodhu listening and asking questions back avlovaa ilaama sometimes nee irundhaa..That was something soo basic that you lacked sometimes..When I told you, you tried to immediately correct from the next time!..

All this speaks a lot about your character!

Some people nee azhaga irukka nu un mela crush varudhu nu
solraanga...

But I feel it is the Reverse...

Your Heart is so beautiful that it is Brilliantly Reflecting on your
face..

You are soo beautiful because your heart speaks of such beauty!
"Agathin Azhagu Mugathil Theriyum"..

Thats whyy you are sooo cute even when your face is full of kutty
kutty cute pimples!!!!

Lets talk about Care...Your care for others is not Loud in the conventional sense..

You may not ask saaptiya or other general ways of care..

You show your care in a different way

Rather, your care shines the brightest when someone close to you is going through their lowest

Initially indha semester la, when my face was pei pudicha maari irunchu laa, entire college was asking enna aachu nu..

As a friend, I remember you sending me many many voice

messages telling me to stop suffering silently about something and bring my focus back to studies..

That and how much you wanted to be there for me as a friend andha Kovil ku pogum bhodhu.. Neeyum vara nu sonnadhu(Enakku enna aagudhu nu therinjikka) - are the sweetest things you ever did to me as a friend ..

And I have also listened to madhushree's stories regarding how much you used to give advice and console her when she is feeling down - neenga munaadi semmaya pesitu irundha modhu laam..

So unakku care laam irukku, it will always show in subtle ways and not conventional ways that one might openly see

Also, the amount of efforts you put when you like someone as a person, like madhushree ku andha birthday gifts ku ready pannadhellam chance eh illa and also how much efforts you put into pipe was also astonishing..avlooo thought and effort podraa avloo sweet ah

But ennaku ennaa naa, sometimes avangalukku adhoda value therila..like pipe for instance..nee avloo efforts potturukkaa..but Avanukku adhu therila..and madhushree bday gift ellam thooki

pottutaa sonaa, Enakku adhellam ketkum bhodhu rombaa bad ah
feel aachu.."cha iva ivlo sweet ah ellam panraa but yen ivala
elaarum ipdi treat panraangaa"...

I am sorry but as a human being I felt like "Cha naan laam
irundhurundhaa chinnadhaa kannu kalangaama princess maari
treat pannirrpen iva ivloo panradhukku " apdinu I felt ..

You out of everyone deserve that level of treatment in your life...

Nee sollittey irukka nee avloo nulla ponnu illa nu...let's revisit
some facts...

Nee poi solla maatengura..maybe you ll hide a few things but poi
solli naan paathadhilla .. kutty kutty poi samaalichification nee

pannirka naan paathukken..but naan solradhu serious
vishayangal la you have never ever lied..

And also while you were in a relationship, you didn't even
thirumbi paathufy other boys who were trying to impress
you...You were exceedingly loyal to pipe..

And most importantly, I can say someone can really trust you
completely and I trust you entirely

These qualities - Honesty, Trustworthy and loyalty - do you know
how rare these three qualities are in this generation?..

Do you know how many stories I have heard and people didn't even have one of these qualities..how can someone not see how special this is...

I would like to give the "Worlds Biggest..." Sorry wait...
The Biggest Idiot of the Universe award goes to Mr.Pipe..
Epdi daa ipdi patta ponnah vittu matha ponukkelaam ponaan is something I will never able to comprehend in this lifetime

Vijay photo vaam Surya photo vaam , vera vera yaam - serupaalayee adikkanum XD

See i understand..unkittayum flaws irukku...but ivlo sweetness and positive qualities irukkum bhodhu , the boyfriend should constructively sit together and sort out together and grow together...I know unakkum mood swings irukku..Unakkum irrational ah kovam varum..Neeyum neraiyya sanda poduva..I fully admire!..Romba soft ah irundhaa life bore adichurum haha!..konjam sandai and spicyness everyone needs lolll..

Ketaachu ketaachu, un chandramukhi face , toxic face elaam ketaachu - idhu sakthi priyaa vaa nu shock aagiyaachu adhellam kettu, ellam December laye mudinjuruchu.. so stop saying Enakku innum therila anol... podhummm 😂 🎉

You asked me what I learnt from you...

I have learnt how exactly I should apologise and with what intensity to someone I deeply care about

I have learnt how to be more considerate about what other people might feel, like avangala bad ah feel panna vekka koodadhu

I have learnt to be more devotional and spiritual

I have learnt how to see the beauty in sharing even the most smallest of things to someone as a friend.. like andha updating

and how beautiful it truly is!

I was already like that in 12th grade but after coming to VIT, i got
soo indulged in the YouTube mindset, that I forgot that part of me
also existed..

You made me recall that version of me three years ago

You introduced me to soo many positive and good things - like
swami songs i didn't know about and soo much more

Haan and yes - Tea ah maranteneyyy 😂 😊

I just feel like I want to learn and grow along with you..

I have overall ah learnt about every single side of you - happy, sad, depressed, overthinking, traumatic, funn side , dark side and even Toxic side, everything

Don't say there is much much more..Overall ah ellam therinjiruchu ennaku and still fell for you..

There is a saying - Comparing yourself to others is an insult to who you are...Andha maari bottle pannadhu, or guhan sonnadhellam please don't compare with me and what I am typing here..

You know how genuine I am and how dheiveegamaa I sincerely admire you! I request you to please listen to that voice message i sent a few days back once again..Even though you fully listened to it, it will make much much more sense now..

I was never going to say all this at this time..
but again, you said you liked something guhan said or did and i couldn't prolong this any further.. That was my signal to immediately say all this..

Enaku honestly light ah Avan overtake panra maari irunchu so i wanted to tell everything before anything else happened to be clear and be genuine on my side

I don't know what kind of drugs i was smoking in the first year to not try all this for an attempt at that point.. ipdi oru ponnah how i missed nu I have absolutely no clue..

I just have one small request from my side, Unnoda adamm or pidivaadham, whatever it is, just oru vaati mattum, orae oru vaati mattum, Enakku oru consideration kudu.. You already like few qualities in me as a friend.. The friendship is indeed special..but Nee consideration kooda kudukka maaten ndradhu is a little tooo harsh for me

Lots of people around the world fell for their bestfriends around the world and are ultimately happy...I am not saying just because avangaluku andha maari success aachu nu namakkum aagum nu solla varala maa..

But nee romba harsh ah idhu nadakaave nadakaadhu and enaku failure dhaan nu solra and mind la fixeddd ah vechurka

I am just saying apdiyum successful relationships irukku so just give me a small consideration

Even if there are 100 people around me, you ll always be running somewhere in my head (in a sweet way - padhara vendaam)

I ll still search for you and try to talk to you and involve you as well..

Importantly, If Mana porutham is there, everything else will get sorted out with patience..

Nee sonna dialogue dhaan..so yes sollitten..so konjam ivlo strict ah ilaama indha pacha mannu mela paavam paarunga mammmm!!!

Enna maari En kitta irukka sila qualities la paiyan venum nu solraaa, why are youu going for duplicate when original is here 😭😭 what do you have soo much against mee(summa

fun ku indha line)

Butt waitt..I totally get it..Unakku andha "feel" varala nu solra...yaarayum we cannot force andha feel vara vekkaa... I am mature enough to understand that..

but yen point enna naa oru consideration and time nee kuduthaaa dhaan andha feel varudhaa illaya nu theriyum...this is true... just give it a little bit of time and consideration in the background of your head is all I am asking for..Don't have to idhuku nu utkaandhu do that, background la vechitu namma velaiya paakalam

Now...Very Important -

What exactly do I mean by consideration?..Just indha maari enaku un mela genuine interest irukku nu , college mudiyura verkkum, background la oru thought la vechiko, that's it, that's all I ask for..just oru small thought...it will make your mind to automatically make a decision in the end..

Just andha time mattum background la kudu

Adhu kaprm I promise I ll accept whatever decision it is that you are going to take and I ll completely understand!

Let the candidate spend some time with you and perform and

then adhu kaprm kooda set aavadhu nu thonuchu naa, you can kick out the candidate

But andha point verkkum Ipdi patta oru ponnuukku i don't mind getting how many more bulbs and clownism
I am alright and rejection is my old friend..I know how to pakkavaa handle it

I don't even have expect you to process all this.. I don't expect you to reply to any of this as well or decide anything. This time period is very stressful since every hour has to be utilised effectively for placements.

Always be 1000% transparent with me and please never hide anything about me regarding any of this...open ah pattu pattu nu pottu ipo naan panra maari udachuru...honesty will smoothen out everything..

Just be a little soft while conveying whatever you want..kutty request adhu...

It is impossible for me to hate you or form any sort of negative feelings towards you in all jenmams, remember!

I just feel a special connection with you and maybe someday you'll realise that as well..Pannala naalum isokayyyy!!..Really!

In the end, i am always okay with whatever decision you take and just want you to be happyyy alwaysss! And that's what I pray in Kovil - let her be truly happy with whomsoever it is...

I ll leave the rest to God..

You said unakku light ah pressurized ah feel aachu yesterday, Suppose unakku ennaku consideration kooda kudukka venaam nu thonuchu naa, it's okayy, vitturu , I don't wanna pressure you, be free, be jollyyy, don't worry about meee

My aim and existence in life is to not raise anyone's bp or put pressure on anyone..Nee consideration kooda enaku thara thonala nu sonaalum I ll happily accept it! Nee kettaaa adha kooda vittu kuduthuruven! Thats just how much I love you! I ll let go and you be free without any pressure or bp!
Hey naa idhellam light hearted ah honest ah unna full ah understand pannikitu solren haha

Finallyy, Ivlo naal epdi pesunomoo apdiye let's continue if you wish..

If you want some space and time, take how much ever you want to process this entire thing and I promise with all my heart that I

won't have or develop any sort of hope and I will not get hurt..idha pathi nethu anupcha WhatsApp messages la podhum podhum ndra alavuku clarification kuduthutten..please read those once if you have doubts reg hope or me getting hurt

I hope after reading all this you realized oruthanga just exist aagitu and dhoorathulu lerundhu or pakkathulerundhu observe pannaaley, if it captivates the heart, love vandhurum.. Nee oruthanga kitta regular contact, texts, or pesradhukum idhu grow aaguradhukkum samandhamey illa..Even if I observed all of this from behind you from a distance and with no contact with you, I would have fallen..idhaan unmai...one side love perumaale

I fell for your soul and existence..Just remember and trust me on this!

Anyways all of this I wanted to just say about you...

If only you saw yourself through my eyes, you would realise how beautiful and magical you truly are!

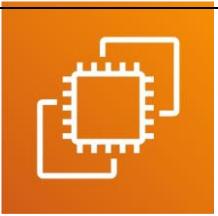
I got your six in every situation ..

I just feel like i want to be on your 3 and your 9 as well..

Vignayyshhh,
Singing off

Amazon 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

EC2 User Data

- It is possible to bootstrap our instances using an [EC2 User data](#) script.
- [bootstrapping](#) means launching commands when a machine starts
- That script is [only run once](#) at the instance [first start](#)
- EC2 user data is used to automate boot tasks such as:
 - Installing updates
 - Installing software
 - Downloading common files from the internet
 - Anything you can think of
- The EC2 User Data Script runs with the root user

Hands-On: Launching an EC2 Instance running Linux

- We'll be launching our first virtual server using the AWS Console
- We'll get a first high-level approach to the various parameters
- We'll see that our web server is launched using EC2 user data
- We'll learn how to start / stop / terminate our instance.

EC2 Instance Types - Overview

- You can use different types of EC2 instances that are optimised for different use cases (<https://aws.amazon.com/ec2/instance-types/>)
- AWS has the following naming convention:

m5.2xlarge

General Purpose

Compute Optimized

Memory Optimized

Accelerated Computing

Storage Optimized

HPC Optimized

Instance Features

Measuring Instance Performance

- m: instance class
- 5: generation (AWS improves them over time)
- 2xlarge: size within the instance class

EC2 Instance Types - General Purpose

- Great for a diversity of workloads such as web servers or code repositories
- Balance between:
 - Compute
 - Memory
 - Networking
- In the course, we will be using the t2.micro which is a General Purpose EC2 instance

General Purpose

General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads. These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories.

Mac	T4g	T3	T3a	T2	M6g	M5	M5a	M5n	M5zn	M4	A1
-----	-----	----	-----	----	-----	----	-----	-----	------	----	----

* this list will evolve over time, please check the AWS website for the latest information

EC2 Instance Types - Compute Optimized

- Great for compute-intensive tasks that require high performance processors:
 - Batch processing workloads
 - Media transcoding
 - High performance web servers
 - High performance computing (HPC)
 - Scientific modeling & machine learning
 - Dedicated gaming servers

Compute Optimized

Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors. Instances belonging to this family are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.

C6g C6gn C5 C5a C5n C4

* this list will evolve over time, please check the AWS website for the latest information

EC2 Instance Types - Memory Optimized

- Fast performance for workloads that process large data sets in memory
- Use cases:
 - High performance, relational/non-relational databases
 - Distributed web scale cache stores
 - In-memory databases optimized for BI (business intelligence)
 - Applications performing real-time processing of big unstructured data

Memory Optimized

Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.

R6g

R5

R5a

R5b

R5n

R4

X1e

X1

High Memory

z1d

* this list will evolve over time, please check the AWS website for the latest information

EC2 Instance Types - Storage Optimized

- Great for storage-intensive tasks that require high, sequential read and write access to large data sets on local storage
- Use cases:
 - High frequency online transaction processing (OLTP) systems
 - Relational & NoSQL databases
 - Cache for in-memory databases (for example, Redis)
 - Data warehousing applications
 - Distributed file systems

Storage Optimized

Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.

I3	I3en	D2	D3	D3en	H1
----	------	----	----	------	----

* this list will evolve over time, please check the AWS website for the latest information

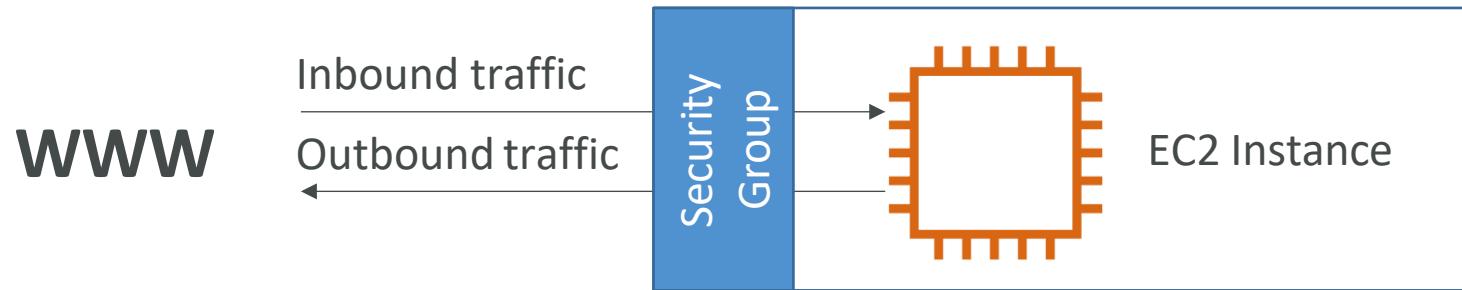
EC2 Instance Types: example

Instance	vCPU	Mem (GiB)	Storage	Network Performance	EBS Bandwidth (Mbps)
t2.micro	1	1	EBS-Only	Low to Moderate	
t2.xlarge	4	16	EBS-Only	Moderate	
c5d.4xlarge	16	32	1 x 400 NVMe SSD	Up to 10 Gbps	4,750
r5.16xlarge	64	512	EBS Only	20 Gbps	13,600
m5.8xlarge	32	128	EBS Only	10 Gbps	6,800

Great website: <https://instances.vantage.sh>

Introduction to Security Groups

- Security Groups are the fundamental of network security in AWS
- They control how traffic is allowed into or out of our EC2 Instances.



- Security groups only contain **allow** rules
- Security groups rules can reference by IP or by security group

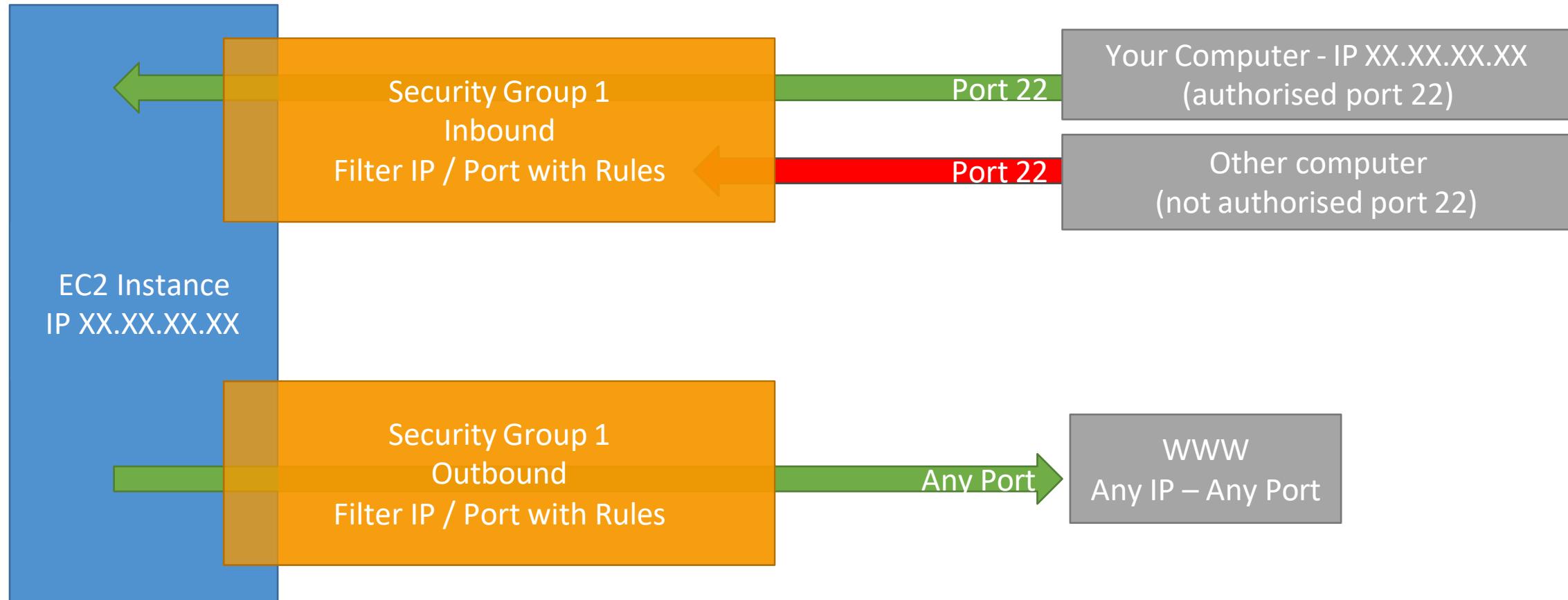
Security Groups

Deeper Dive

- Security groups are acting as a “firewall” on EC2 instances
- They regulate:
 - Access to Ports
 - Authorised IP ranges - IPv4 and IPv6
 - Control of inbound network (from other to the instance)
 - Control of outbound network (from the instance to other)

Type	Protocol	Port Range	Source	Description
HTTP	TCP	80	0.0.0.0/0	test http page
SSH	TCP	22	122.149.196.85/32	
Custom TCP Rule	TCP	4567	0.0.0.0/0	java app

Security Groups Diagram



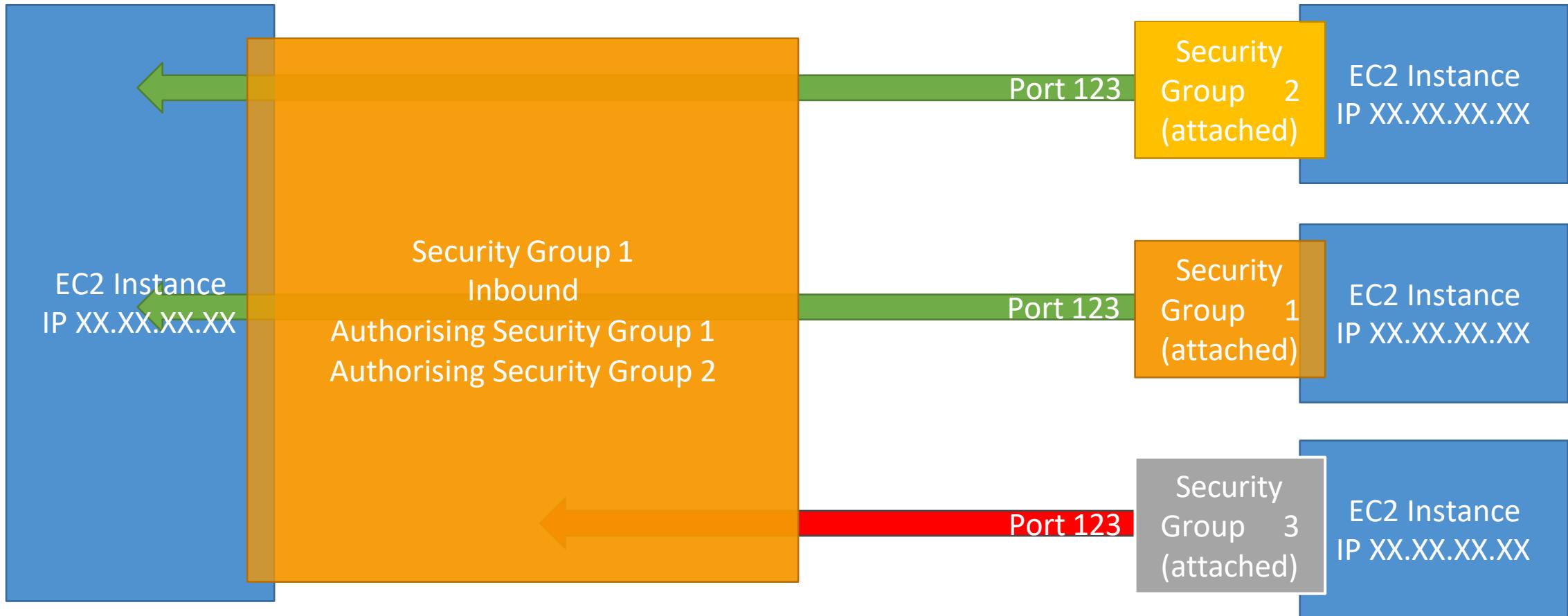
Security Groups

Good to know

- Can be attached to multiple instances
- Locked down to a region / VPC combination
- Does live “outside” the EC2 - if traffic is blocked the EC2 instance won’t see it
- **It’s good to maintain one separate security group for SSH access**
- If your application is not accessible (time out), then it’s a security group issue
- If your application gives a “connection refused“ error, then it’s an application error or it’s not launched
- All inbound traffic is **blocked** by default
- All outbound traffic is **authorised** by default

Referencing other security groups

Diagram



Classic Ports to know

- 22 = SSH (Secure Shell) - log into a Linux instance
- 21 = FTP (File Transfer Protocol) - upload files into a file share
- 22 = SFTP (Secure File Transfer Protocol) - upload files using SSH
- 80 = HTTP - access unsecured websites
- 443 = HTTPS - access secured websites
- 3389 = RDP (Remote Desktop Protocol) - log into a Windows instance

SSH Summary Table

	SSH	Putty	EC2 Instance Connect
Mac	✓		✓
Linux	✓		✓
Windows < 10		✓	✓
Windows >= 10	✓	✓	✓

Which Lectures to watch

- Mac / Linux:
 - SSH on Mac/Linux lecture
- Windows:
 - Putty Lecture
 - If Windows 10: SSH on Windows 10 lecture
- All:
 - EC2 Instance Connect lecture

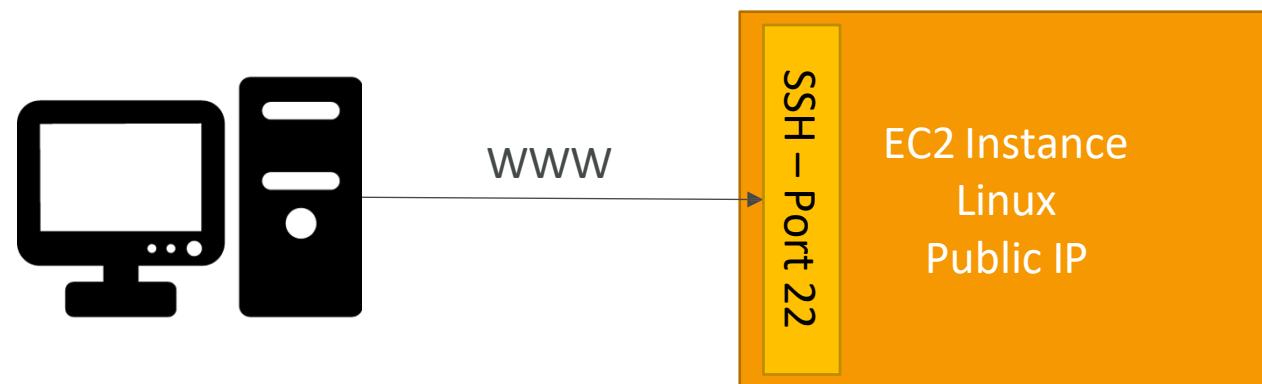
SSH troubleshooting

- Students have the most problems with SSH
- If things don't work...
 1. Re-watch the lecture. You may have missed something
 2. Read the troubleshooting guide
 3. Try EC2 Instance Connect
- If one method works (SSH, Putty or EC2 Instance Connect) you're good
- If no method works, that's okay, the course won't use SSH much

How to SSH into your EC2 Instance

Linux / Mac OS X

- We'll learn how to SSH into your EC2 instance using [Linux / Mac](#)
- SSH is one of the most important function. It allows you to control a remote machine, all using the command line.

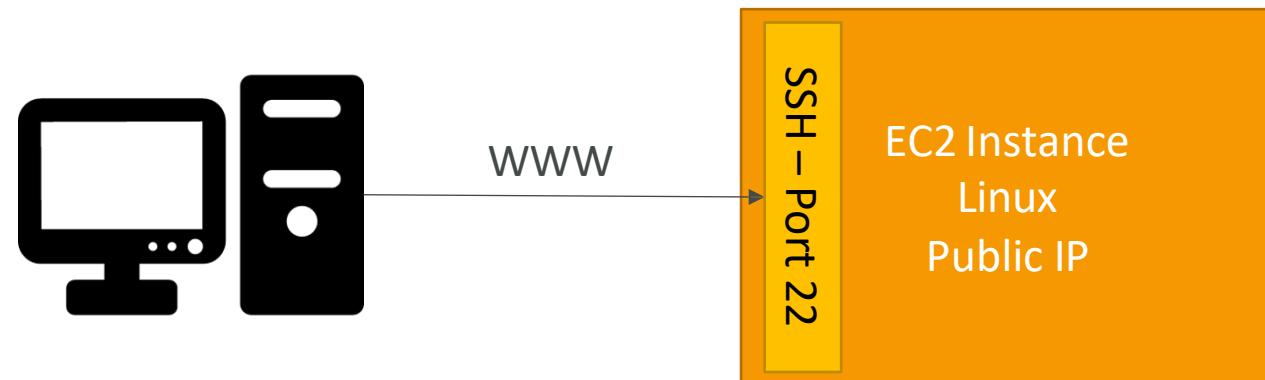


- We will see how we can configure OpenSSH [~/.ssh/config](#) to facilitate the SSH into our EC2 instances

How to SSH into your EC2 Instance

Windows

- We'll learn how to SSH into your EC2 instance using [Windows](#)
- SSH is one of the most important function. It allows you to control a remote machine, all using the command line.



- We will configure all the required parameters necessary for doing SSH on Windows using the free tool [Putty](#).

EC2 Instance Connect

- Connect to your EC2 instance within your browser
- No need to use your key file that was downloaded
- The “magic” is that a temporary key is uploaded onto EC2 by AWS
- Works only out-of-the-box with Amazon Linux 2
- Need to make sure the port 22 is still opened!

EC2 Instances Purchasing Options

- On-Demand Instances - short workload, predictable pricing, pay by second
- Reserved (1 & 3 years)
 - Reserved Instances - long workloads
 - Convertible Reserved Instances - long workloads with flexible instances
- Savings Plans (1 & 3 years) -commitment to an amount of usage, long workload
- Spot Instances - short workloads, cheap, can lose instances (less reliable)
- Dedicated Hosts - book an entire physical server, control instance placement
- Dedicated Instances - no other customers will share your hardware
- Capacity Reservations - reserve capacity in a specific AZ for any duration

EC2 On Demand

- Pay for what you use:
 - Linux or Windows - billing per second, after the first minute
 - All other operating systems - billing per hour
- Has the highest cost but no upfront payment
- No long-term commitment
- Recommended for short-term and un-interrupted workloads, where you can't predict how the application will behave

EC2 Reserved Instances

- Up to **72%** discount compared to On-demand
- You reserve a specific instance attributes (Instance Type, Region, Tenancy, OS)
- Reservation Period - 1 year (+discount) or 3 years (+++discount)
- Payment Options - No Upfront (+), Partial Upfront (++) , All Upfront (+++)
- Reserved Instance's Scope - Regional or Zonal (reserve capacity in an AZ)
- Recommended for steady-state usage applications (think database)
- You can buy and sell in the Reserved Instance Marketplace
- Convertible Reserved Instance
 - Can change the EC2 instance type, instance family, OS, scope and tenancy
 - Up to **66%** discount

Note: the % discounts are different from the video as AWS change them over time – the exact numbers are not needed for the exam. This is just for illustrative purposes ☺

EC2 Savings Plans

- Get a discount based on long-term usage (up to 72% - same as RIs)
- Commit to a certain type of usage (\$10/hour for 1 or 3 years)
- Usage beyond EC2 Savings Plans is billed at the On-Demand price
- Locked to a specific instance family & AWS region (e.g., M5 in us-east-1)
- Flexible across:
 - Instance Size (e.g., m5.xlarge, m5.2xlarge)
 - OS (e.g., Linux, Windows)
 - Tenancy (Host, Dedicated, Default)



EC2 Spot Instances

- Can get a discount of up to 90% compared to On-demand
- Instances that you can “lose” at any point of time if your max price is less than the current spot price
- The MOST cost-efficient instances in AWS
- Useful for workloads that are resilient to failure
 - Batch jobs
 - Data analysis
 - Image processing
 - Any distributed workloads
 - Workloads with a flexible start and end time
- Not suitable for critical jobs or databases

EC2 Dedicated Hosts

- A physical server with EC2 instance capacity fully dedicated to your use
- Allows you address compliance requirements and use your existing server-bound software licenses (per-socket, per-core, per-VM software licenses)
- Purchasing Options:
 - On-demand - pay per second for active Dedicated Host
 - Reserved - 1 or 3 years (No Upfront, Partial Upfront, All Upfront)
- The most expensive option
- Useful for software that have complicated licensing model (BYOL - Bring Your Own License)
- Or for companies that have strong regulatory or compliance needs

EC2 Dedicated Instances

- Instances run on hardware that's dedicated to you
- May share hardware with other instances in same account
- No control over instance placement (can move hardware after Stop / Start)

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

EC2 Capacity Reservations

- Reserve On-Demand instances capacity in a specific AZ for any duration
- You always have access to EC2 capacity when you need it
- No time commitment (create/cancel anytime), no billing discounts
- Combine with Regional Reserved Instances and Savings Plans to benefit from billing discounts
- You're charged at On-Demand rate whether you run instances or not
- Suitable for short-term, uninterrupted workloads that needs to be in a specific AZ

Which purchasing option is right for me?



- On demand: coming and staying in resort whenever we like, we pay the full price
- Reserved: like planning ahead and if we plan to stay for a long time, we may get a good discount.
- Savings Plans: pay a certain amount per hour for certain period and stay in any room type (e.g., King, Suite, Sea View, ...)
- Spot instances: the hotel allows people to bid for the empty rooms and the highest bidder keeps the rooms. You can get kicked out at any time
- Dedicated Hosts: We book an entire building of the resort
- Capacity Reservations: you book a room for a period with full price even you don't stay in it

Price Comparison

Example - m4.large - us-east-1

Price Type	Price (per hour)
On-Demand	\$0.10
Spot Instance (Spot Price)	\$0.038 - \$0.039 (up to 61% off)
Reserved Instance (1 year)	\$0.062 (No Upfront) - \$0.058 (All Upfront)
Reserved Instance (3 years)	\$0.043 (No Upfront) - \$0.037 (All Upfront)
EC2 Savings Plan (1 year)	\$0.062 (No Upfront) - \$0.058 (All Upfront)
Reserved Convertible Instance (1 year)	\$0.071 (No Upfront) - \$0.066 (All Upfront)
Dedicated Host	On-Demand Price
Dedicated Host Reservation	Up to 70% off
Capacity Reservations	On-Demand Price

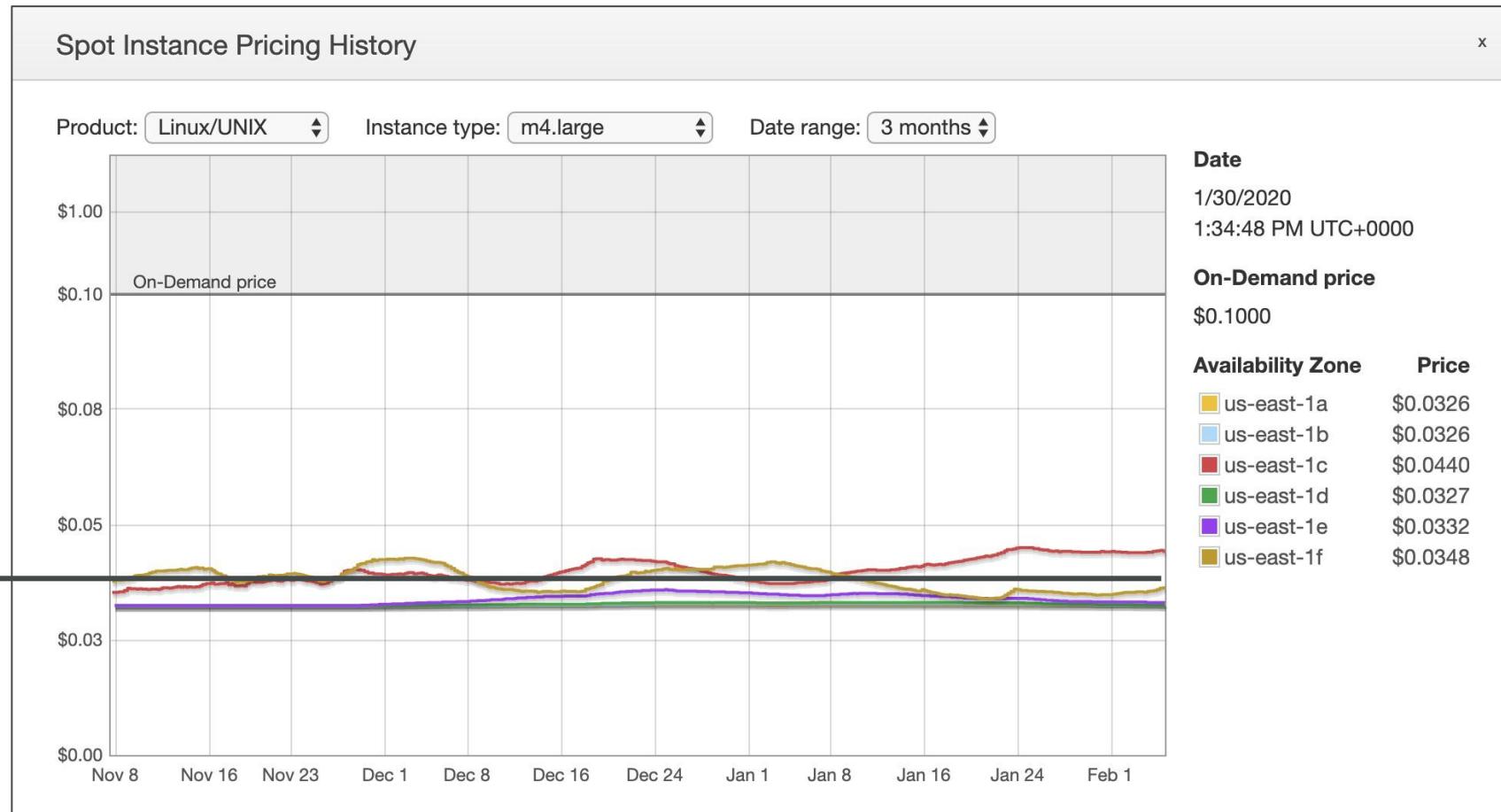


EC2 Spot Instance Requests

- Can get a discount of up to 90% compared to On-demand
- Define max spot price and get the instance while current spot price < max
 - The hourly spot price varies based on offer and capacity
 - If the current spot price > your max price you can choose to stop or terminate your instance with a 2 minutes grace period.
- Other strategy: Spot Block
 - “block” spot instance during a specified time frame (1 to 6 hours) without interruptions
 - In rare situations, the instance may be reclaimed
- Used for batch jobs, data analysis, or workloads that are resilient to failures.
- Not great for critical jobs or databases

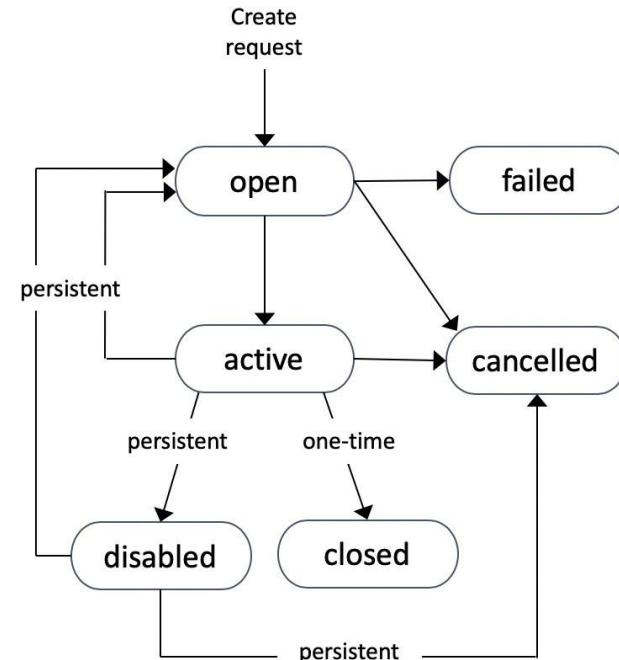
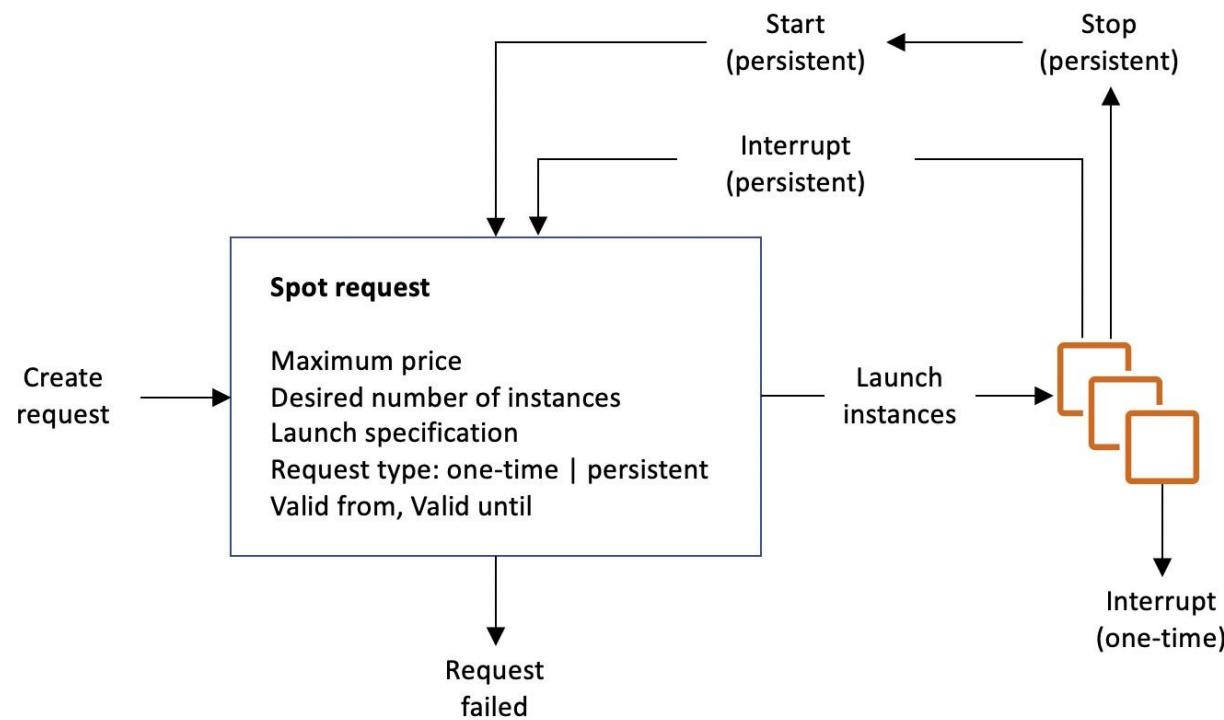
EC2 Spot Instances Pricing

User-defined max price



<https://console.aws.amazon.com/ec2sp/v1/spot/home?region=us-east-1#>

How to terminate Spot Instances?



You can only cancel Spot Instance requests that are **open, active, or disabled**.

Cancelling a Spot Request does not terminate instances

You must first cancel a Spot Request, and then terminate the associated Spot Instances

Spot Fleets

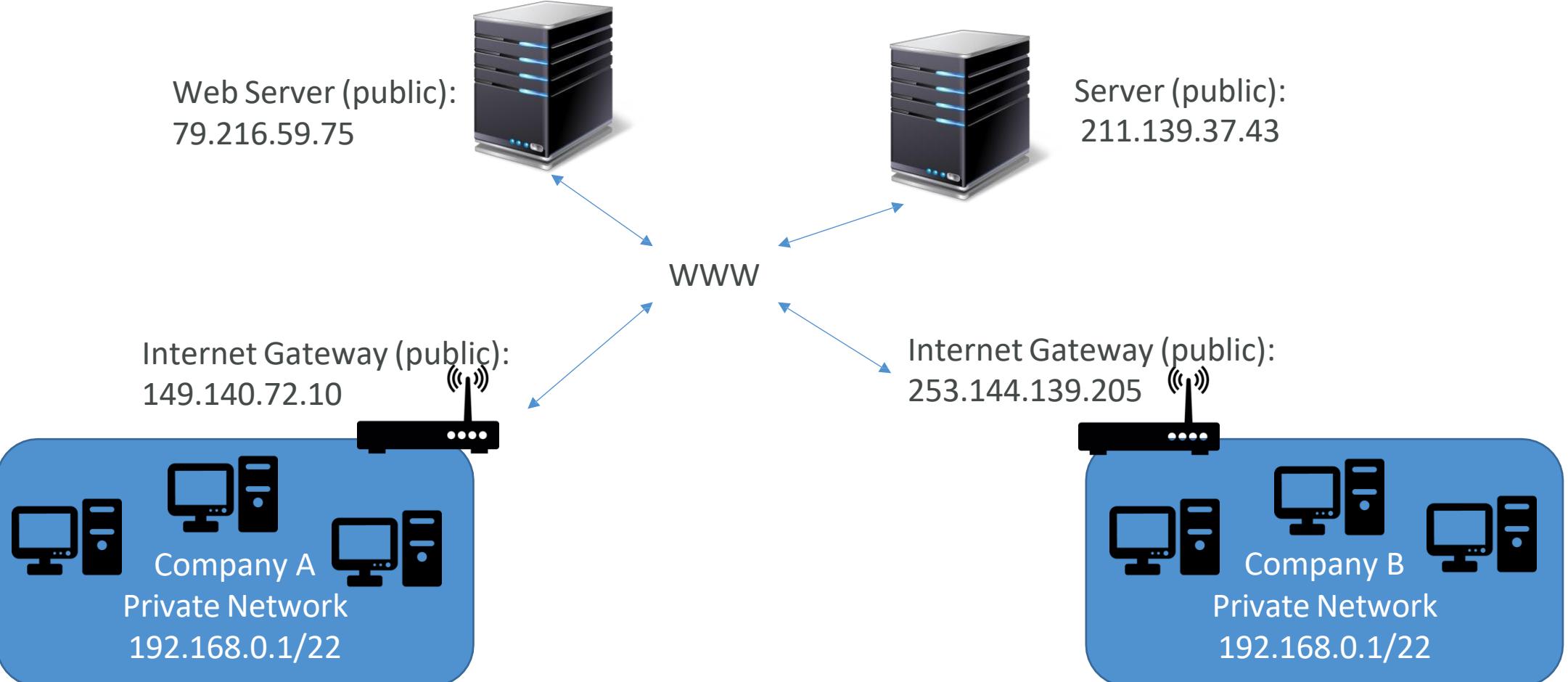
- Spot Fleets = set of Spot Instances + (optional) On-Demand Instances
- The Spot Fleet will try to meet the target capacity with price constraints
 - Define possible launch pools: instance type (m5.large), OS, Availability Zone
 - Can have multiple launch pools, so that the fleet can choose
 - Spot Fleet stops launching instances when reaching capacity or max cost
- Strategies to allocate Spot Instances:
 - lowestPrice: from the pool with the lowest price (cost optimization, short workload)
 - diversified: distributed across all pools (great for availability, long workloads)
 - capacityOptimized: pool with the optimal capacity for the number of instances
 - priceCapacityOptimized (recommended): pools with highest capacity available, then select the pool with the lowest price (best choice for most workloads)
- Spot Fleets allow us to automatically request Spot Instances with the lowest price

Amazon EC2 - Associate

Private vs Public IP (IPv4)

- Networking has two sorts of IPs. IPv4 and IPv6:
 - IPv4: **1.160.10.240**
 - IPv6: **3ffe:1900:4545:3:200:f8ff:fe21:67cf**
- In this course, we will only be using IPv4.
- IPv4 is still the most common format used online.
- IPv6 is newer and solves problems for the Internet of Things (IoT).
- IPv4 allows for **3.7 billion** different addresses in the public space
- IPv4: [0-255].[0-255].[0-255].[0-255].

Private vs Public IP (IPv4) Example



Private vs Public IP (IPv4)

Fundamental Differences

- Public IP:
 - Public IP means the machine can be identified on the internet (WWW)
 - Must be unique across the whole web (not two machines can have the same public IP).
 - Can be geo-located easily
- Private IP:
 - Private IP means the machine can only be identified on a private network only
 - The IP must be unique across the private network
 - BUT two different private networks (two companies) can have the same IPs.
 - Machines connect to WWW using a NAT + internet gateway (a proxy)
 - Only a specified range of IPs can be used as private IP

Elastic IPs

- When you stop and then start an EC2 instance, it can change its public IP.
- If you need to have a fixed public IP for your instance, you need an Elastic IP
- An Elastic IP is a public IPv4 IP you own as long as you don't delete it
- You can attach it to one instance at a time

Elastic IP

- With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.
- You can only have 5 Elastic IP in your account (you can ask AWS to increase that).
- Overall, try to avoid using Elastic IP:
 - They often reflect poor architectural decisions
 - Instead, use a random public IP and register a DNS name to it
 - Or, as we'll see later, use a Load Balancer and don't use a public IP

Private vs Public IP (IPv4)

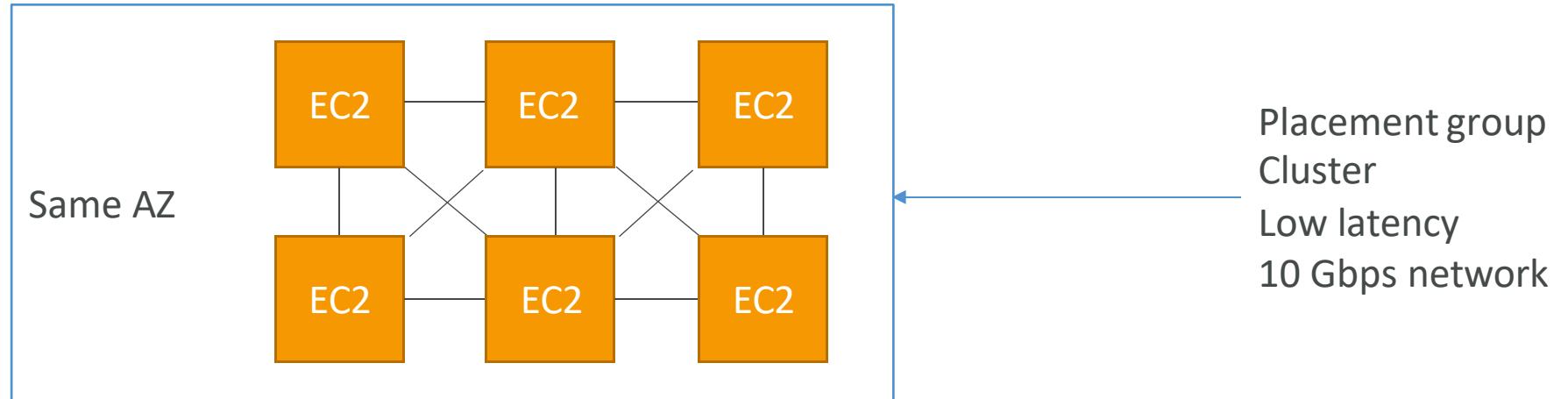
In AWS EC2 - Hands On

- By default, your EC2 machine comes with:
 - A private IP for the internal AWS Network
 - A public IP, for the WWW.
- When we are doing SSH into our EC2 machines:
 - We can't use a private IP, because we are not in the same network
 - We can only use the public IP.
- If your machine is stopped and then started,
the public IP can change

Placement Groups

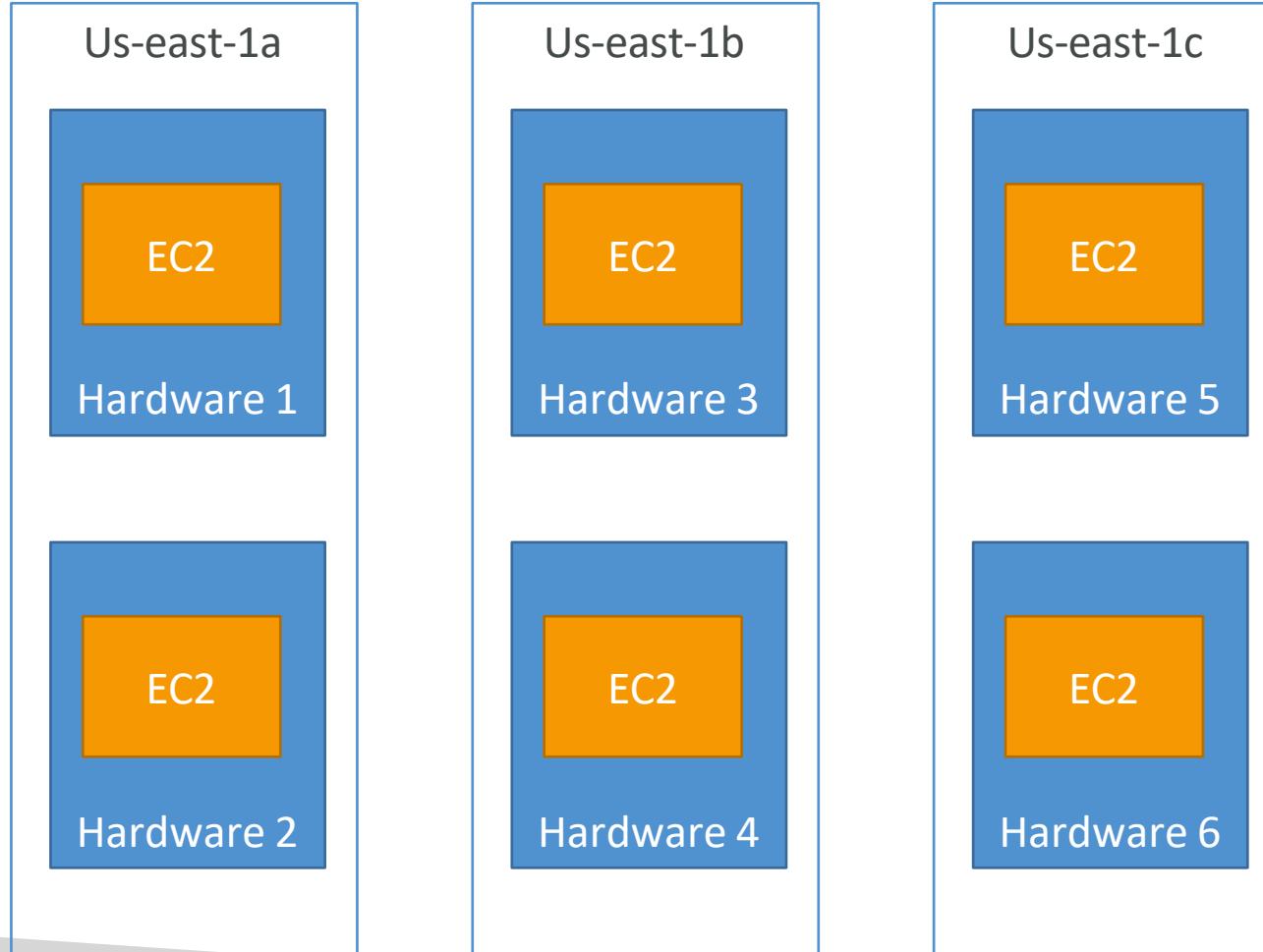
- Sometimes you want control over the EC2 Instance placement strategy
- That strategy can be defined using placement groups
- When you create a placement group, you specify one of the following strategies for the group:
 - *Cluster*—clusters instances into a low-latency group in a single Availability Zone
 - *Spread*—spreads instances across underlying hardware (max 7 instances per group per AZ)
 - *Partition*—spreads instances across many different partitions (which rely on different sets of racks) within an AZ. Scales to 100s of EC2 instances per group (Hadoop, Cassandra, Kafka)

Placement Groups Cluster



- Pros: Great network (10 Gbps bandwidth between instances with Enhanced Networking enabled - recommended)
- Cons: If the AZ fails, all instances fail at the same time
- Use case:
 - Big Data job that needs to complete fast
 - Application that needs extremely low latency and high network throughput

Placement Groups Spread



- Pros:

- Can span across Availability Zones (AZ)
- Reduced risk of simultaneous failure
- EC2 Instances are on different physical hardware

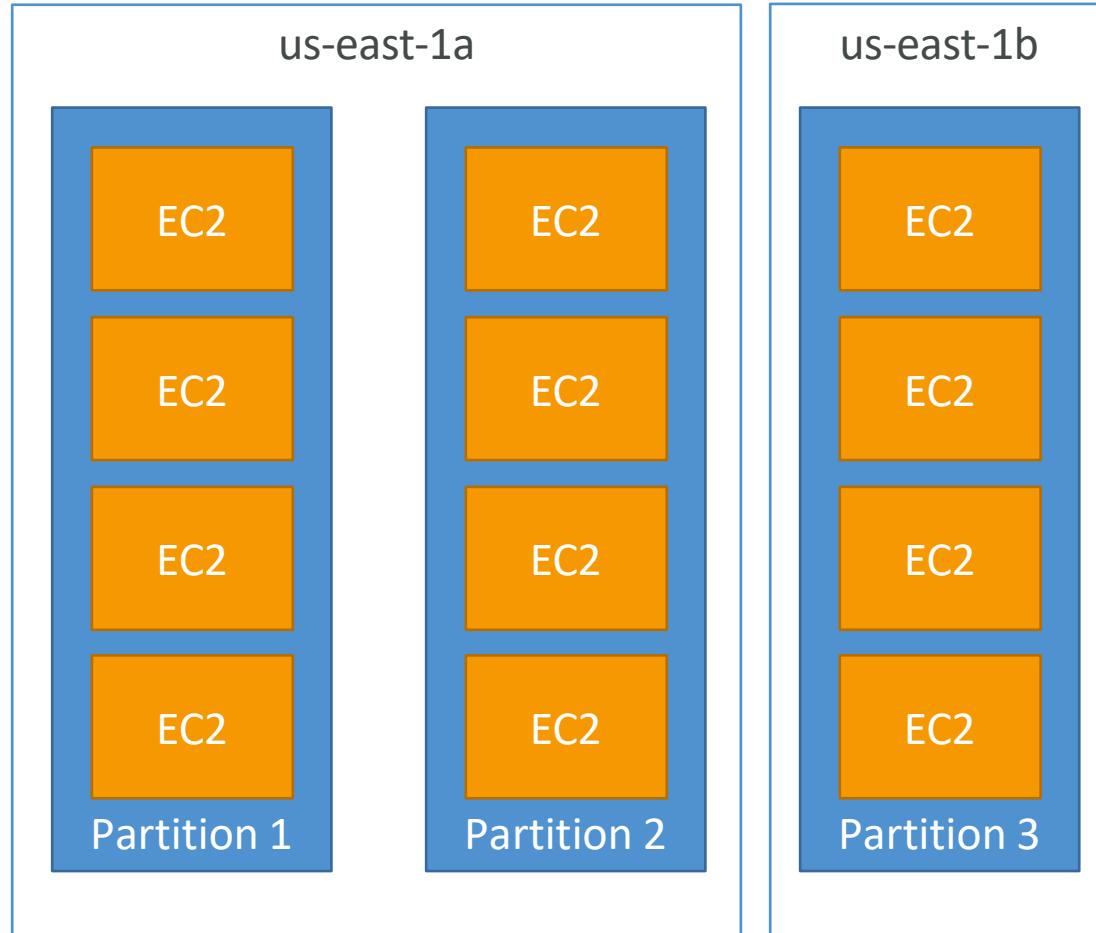
- Cons:

- Limited to 7 instances per AZ per placement group

- Use case:

- Application that needs to maximize high availability
- Critical Applications where each instance must be isolated from failure from each other

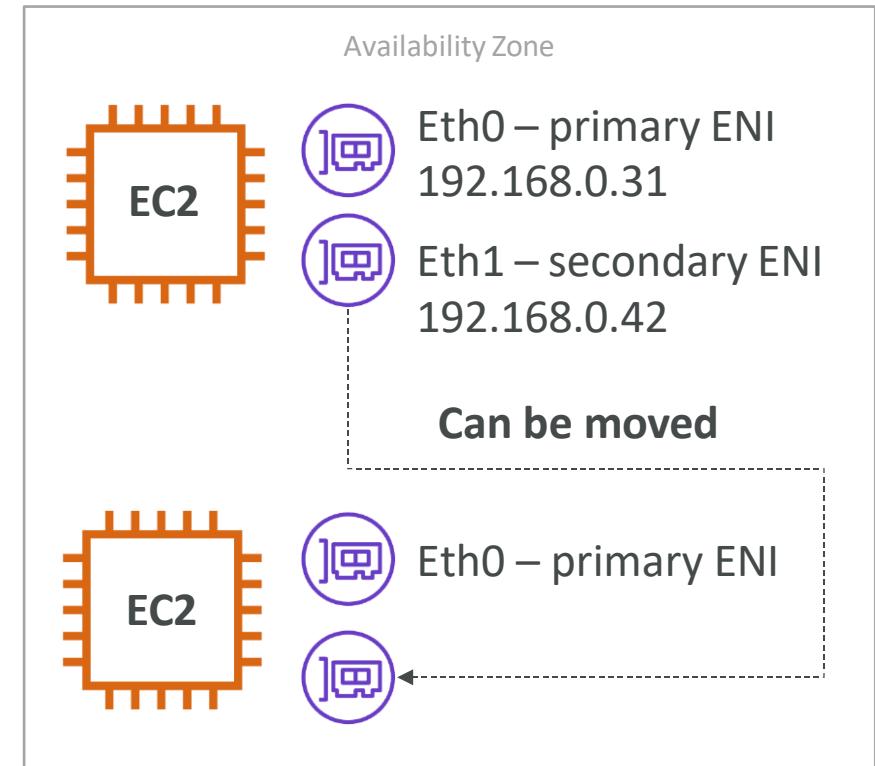
Placements Groups Partition



- Up to 7 partitions per AZ
- Can span across multiple AZs in the same region
- Up to 100s of EC2 instances
- The instances in a partition do not share racks with the instances in the other partitions
- A partition failure can affect many EC2 but won't affect other partitions
- EC2 instances get access to the partition information as metadata
- Use cases: HDFS, HBase, Cassandra, Kafka

Elastic Network Interfaces (ENI)

- Logical component in a VPC that represents a virtual network card
- The ENI can have the following attributes:
 - Primary private IPv4, one or more secondary IPv4
 - One Elastic IP (IPv4) per private IPv4
 - One Public IPv4
 - One or more security groups
 - A MAC address
- You can create ENI independently and attach them on the fly (move them) on EC2 instances for failover
- Bound to a specific availability zone (AZ)

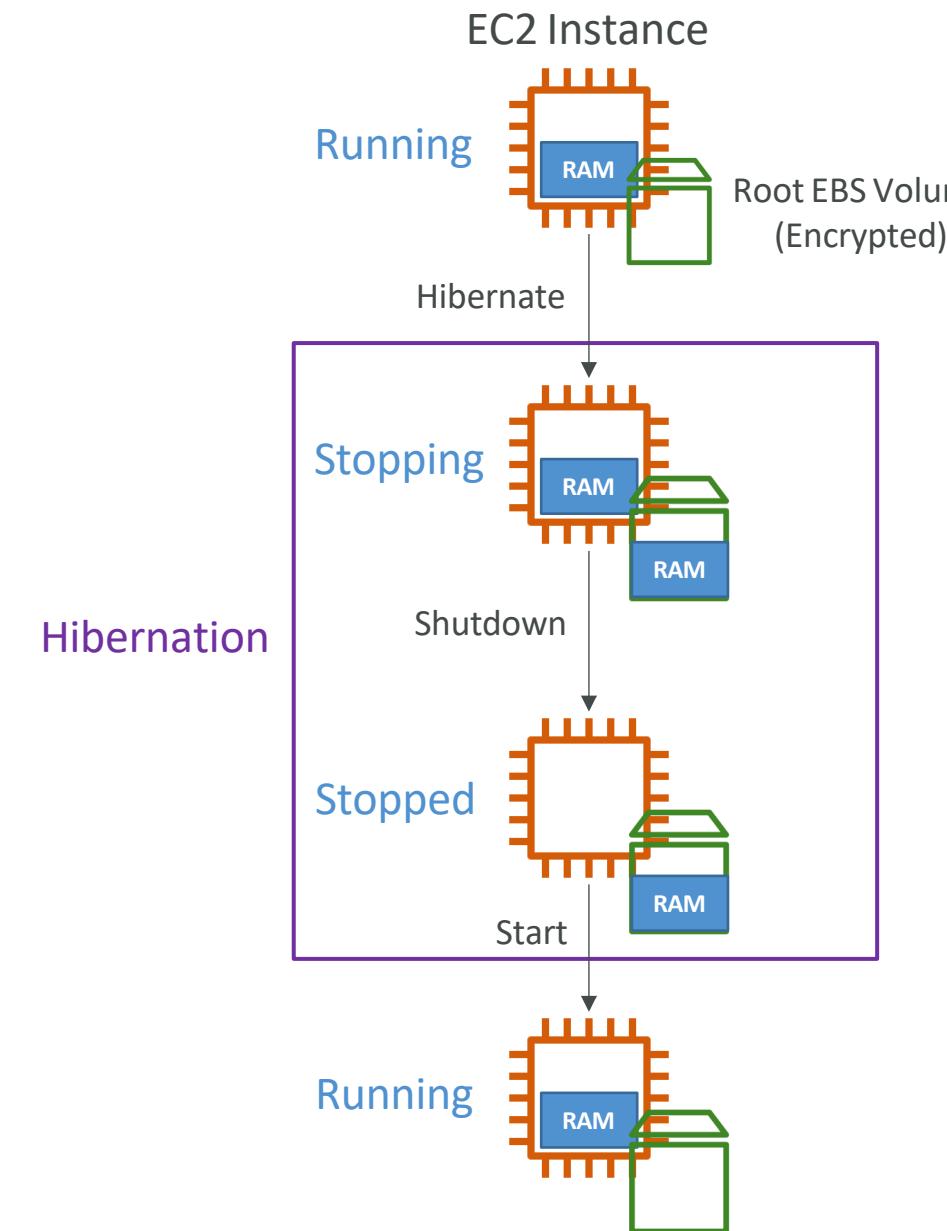


EC2 Hibernate

- We know we can stop, terminate instances
 - Stop - the data on disk (EBS) is kept intact in the next start
 - Terminate - any EBS volumes (root) also set-up to be destroyed is lost
- On start, the following happens:
 - First start: the OS boots & the EC2 User Data script is run
 - Following starts: the OS boots up
 - Then your application starts, caches get warmed up, and that can take time!

EC2 Hibernate

- Introducing EC2 Hibernate:
 - The in-memory (RAM) state is preserved
 - The instance boot is much faster! (the OS is not stopped / restarted)
 - Under the hood: the RAM state is written to a file in the root EBS volume
 - The root EBS volume must be encrypted
- Use cases:
 - Long-running processing
 - Saving the RAM state
 - Services that take time to initialize

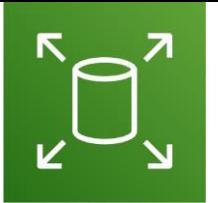


EC2 Hibernate - Good to know

- Supported Instance Families - C3, C4, C5, I3, M3, M4, R3, R4,T2,T3, ...
- Instance RAM Size - must be less than 150 GB.
- Instance Size - not supported for bare metal instances.
- AMI - Amazon Linux 2, Linux AMI, Ubuntu, RHEL, CentOS & Windows...
- Root Volume - must be EBS, encrypted, not instance store, and large
- Available for On-Demand, Reserved and Spot Instances
- An instance can NOT be hibernated more than 60 days

Amazon EC2 - Instance Storage

What's an EBS Volume?

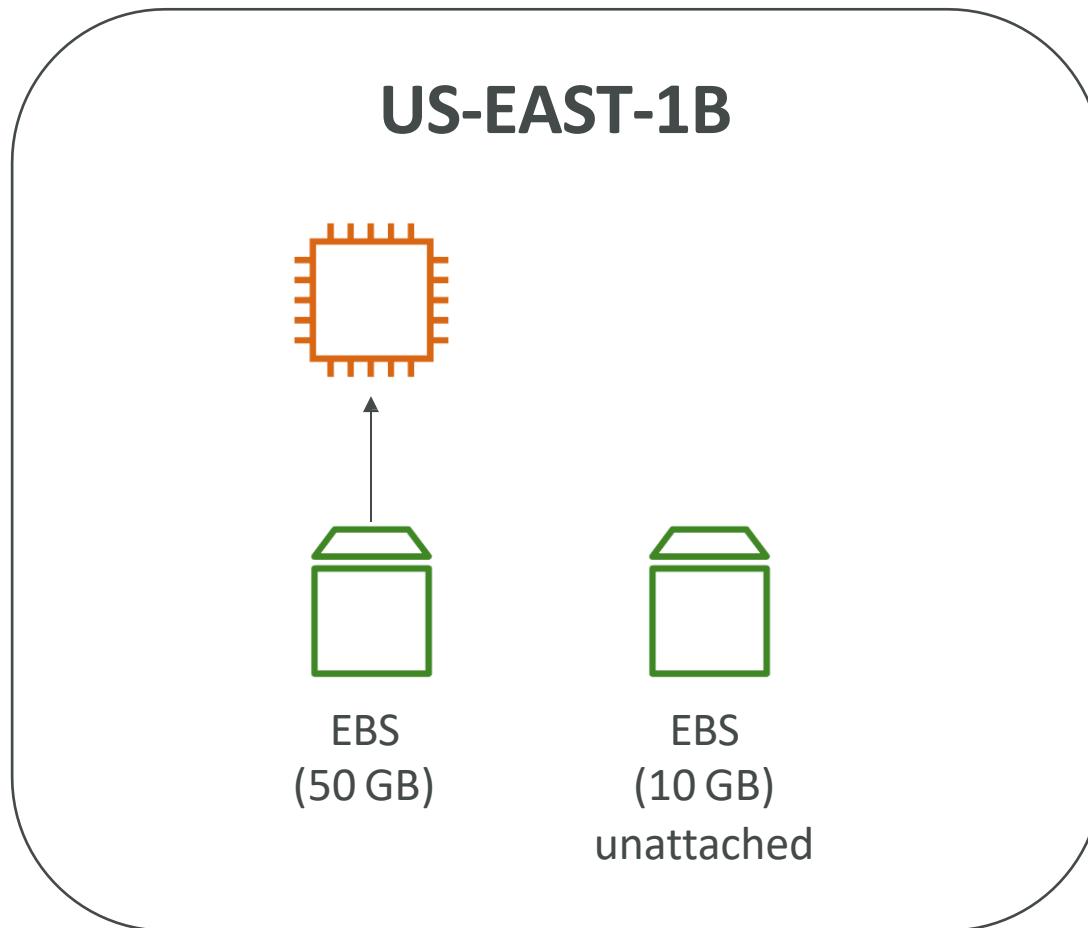
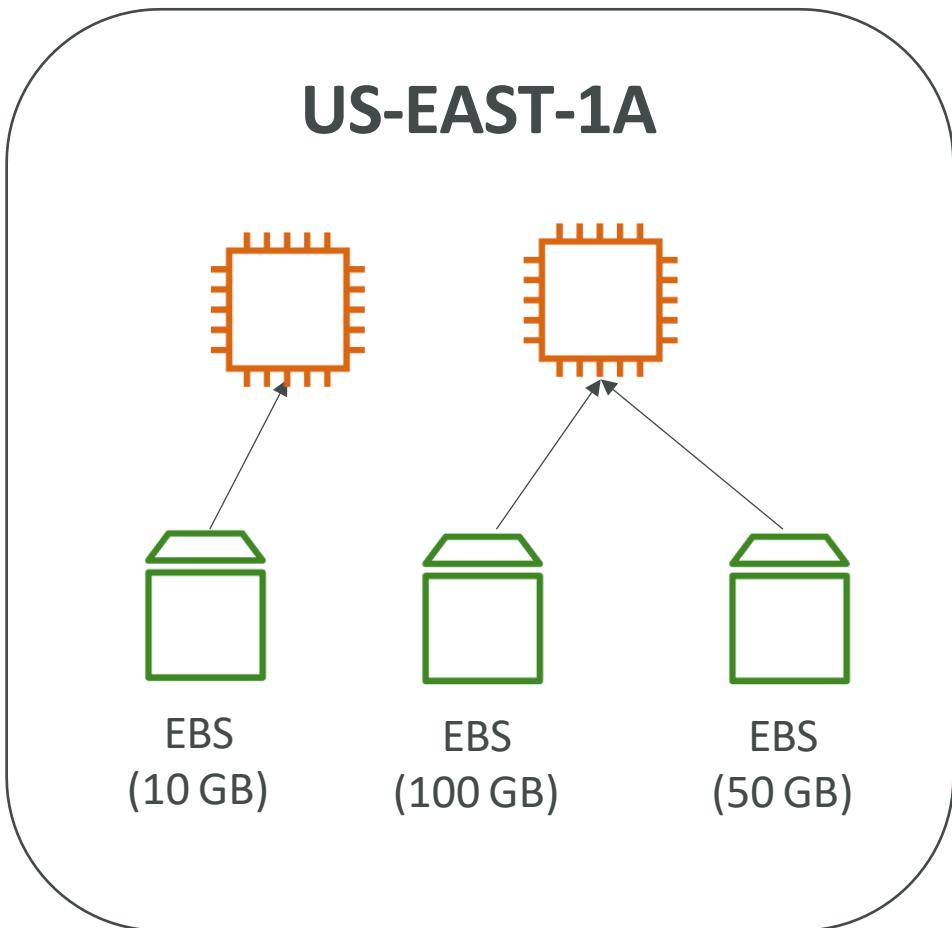


- An **EBS (Elastic Block Store) Volume** is a **network** drive you can attach to your instances while they run
- It allows your instances to persist data, even after their termination
- They can only be mounted to one instance at a time (at the CCP level)
- They are bound to a specific availability zone
- Analogy: Think of them as a “network USB stick”

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



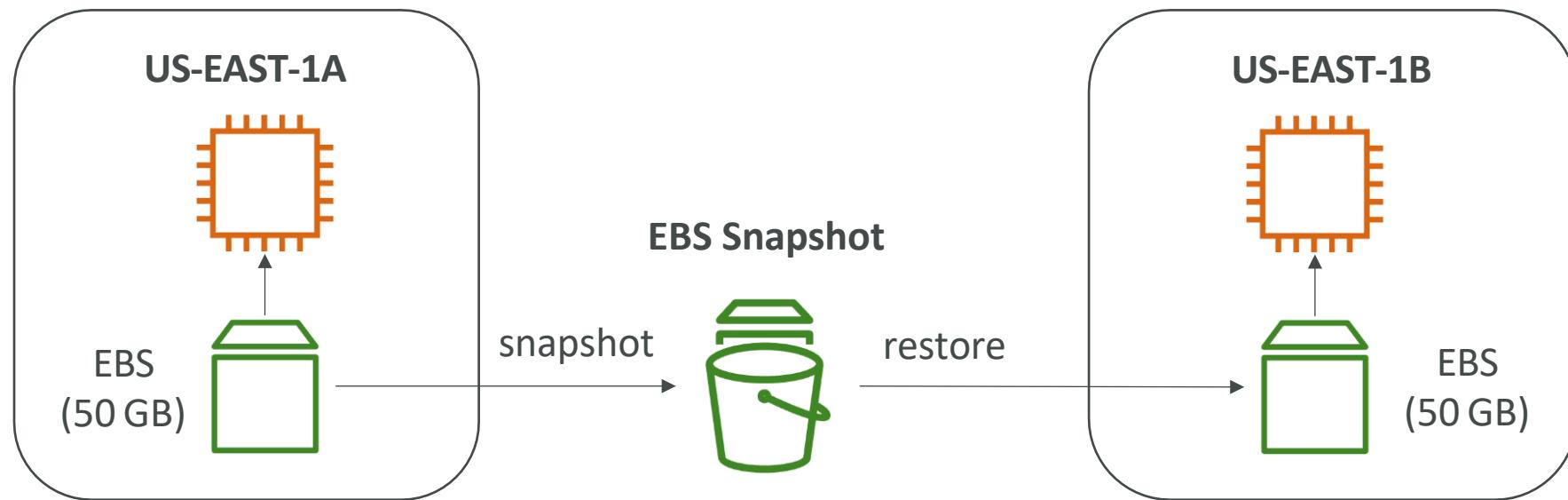
EBS - Delete on Termination attribute

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encryption
Root	/dev/xvda	snap-09f18f682fd23a1b1	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted
EBS	/dev/sdb	Search (case-insensit)	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input type="checkbox"/>	Not Encrypted
Add New Volume								

- Controls the EBS behaviour 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

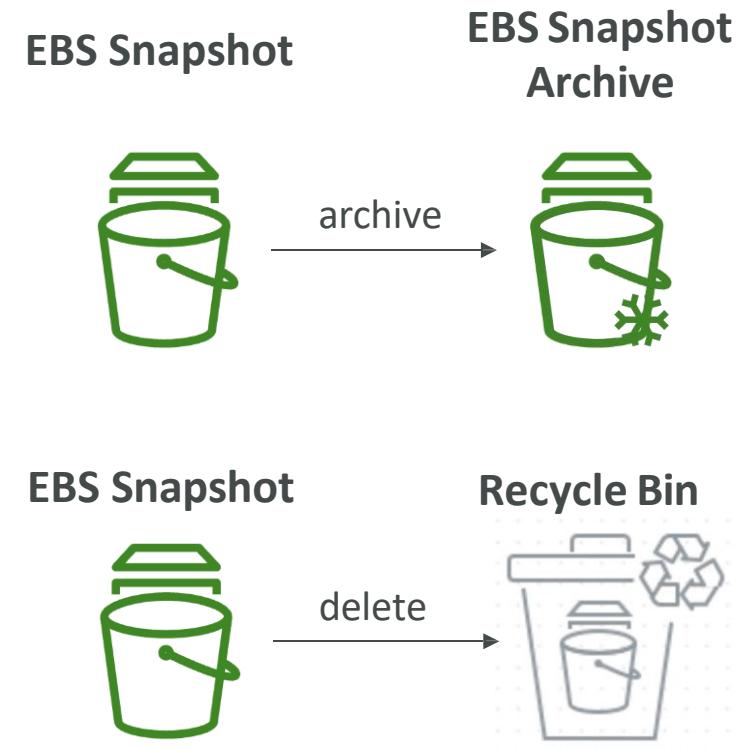
EBS Snapshots

- 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 (\$\$\$)



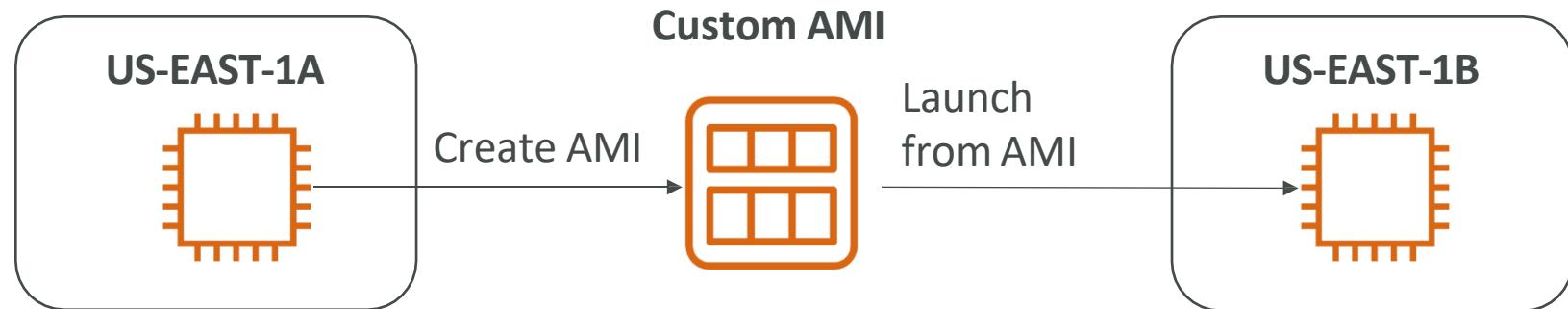
AMI Overview

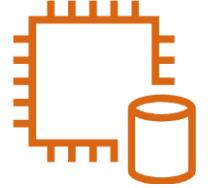


- AMI = Amazon Machine Image
- AMI are a customization of an EC2 instance
 - You add your own software, configuration, operating system, monitoring...
 - Faster boot / configuration time because all your software is pre-packaged
- AMI are built for a specific region (and can be copied across regions)
- 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

Local EC2 Instance Store

Very high IOPS

Instance Size	100% Random Read IOPS	Write IOPS
i3.large *	100,125	35,000
i3.xlarge *	206,250	70,000
i3.2xlarge	412,500	180,000
i3.4xlarge	825,000	360,000
i3.8xlarge	1.65 million	720,000
i3.16xlarge	3.3 million	1.4 million
i3.metal	3.3 million	1.4 million
i3en.large *	42,500	32,500
i3en.xlarge *	85,000	65,000
i3en.2xlarge *	170,000	130,000
i3en.3xlarge	250,000	200,000
i3en.6xlarge	500,000	400,000
i3en.12xlarge	1 million	800,000
i3en.24xlarge	2 million	1.6 million
i3en.metal	2 million	1.6 million

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 Block Express (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 Block Express 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 (4 GiB - 16 TiB):
 - Max PIOPS: 64,000 for Nitro EC2 instances & 32,000 for other
 - Can increase PIOPS independently from storage size
- 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

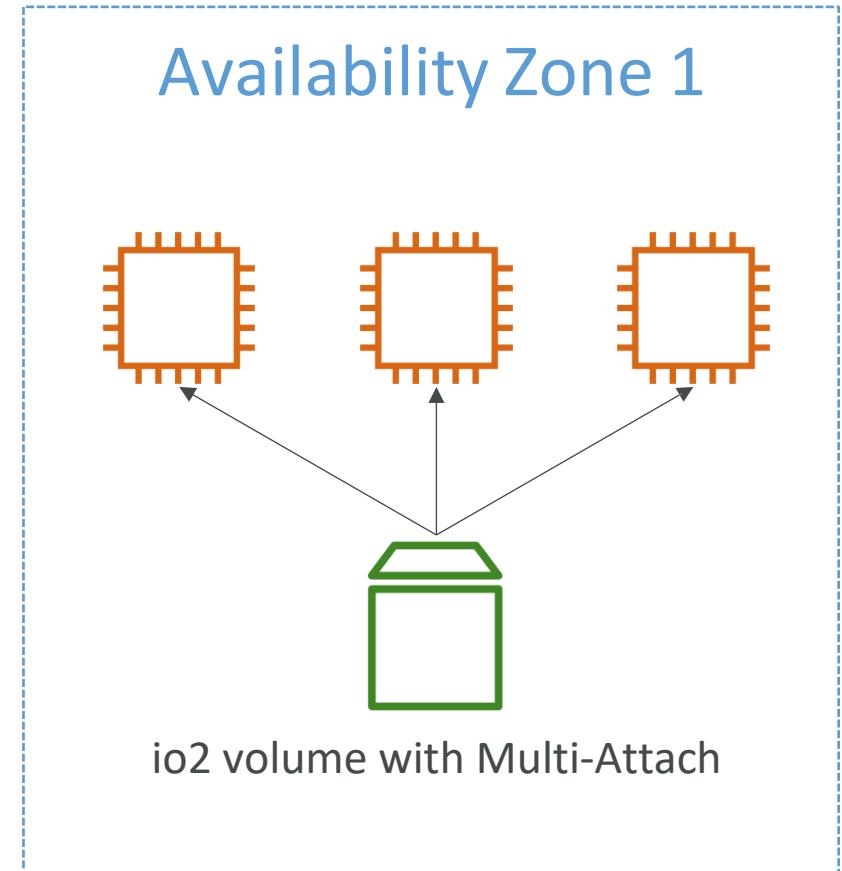
	General Purpose SSD volumes		Provisioned IOPS SSD volumes	
Volume type	gp3	gp2	io2 Block Express ³	io1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.999% durability (0.001% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	<ul style="list-style-type: none"> Transactional workloads Virtual desktops Medium-sized, single-instance databases Low-latency interactive applications Boot volumes Development and test environments 	<p>Workloads that require:</p> <ul style="list-style-type: none"> Sub-millisecond latency Sustained IOPS performance More than 64,000 IOPS or 1,000 MiB/s of throughput 	<ul style="list-style-type: none"> 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	4 GiB - 16 TiB	
Max IOPS per volume (16 KiB I/O)	16,000	256,000 ⁵	64,000	
Max throughput per volume	1,000 MiB/s	250 MiB/s ¹	4,000 MiB/s	1,000 MiB/s ²
Amazon EBS Multi-attach	Not supported		Supported	
NVMe reservations	Not supported		Supported	Not supported
Boot volume	Supported			

	Throughput Optimized HDD volumes	Cold HDD volumes
Volume type	st1	sc1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	
Use cases	<ul style="list-style-type: none"> Big data Data warehouses Log processing 	<ul style="list-style-type: none"> Throughput-oriented storage for data that is infrequently accessed Scenarios where the lowest storage cost is important
Volume size	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	
Boot volume	Not supported	

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#solid-state-drives>

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, EXT4, etc...)



EBS Encryption

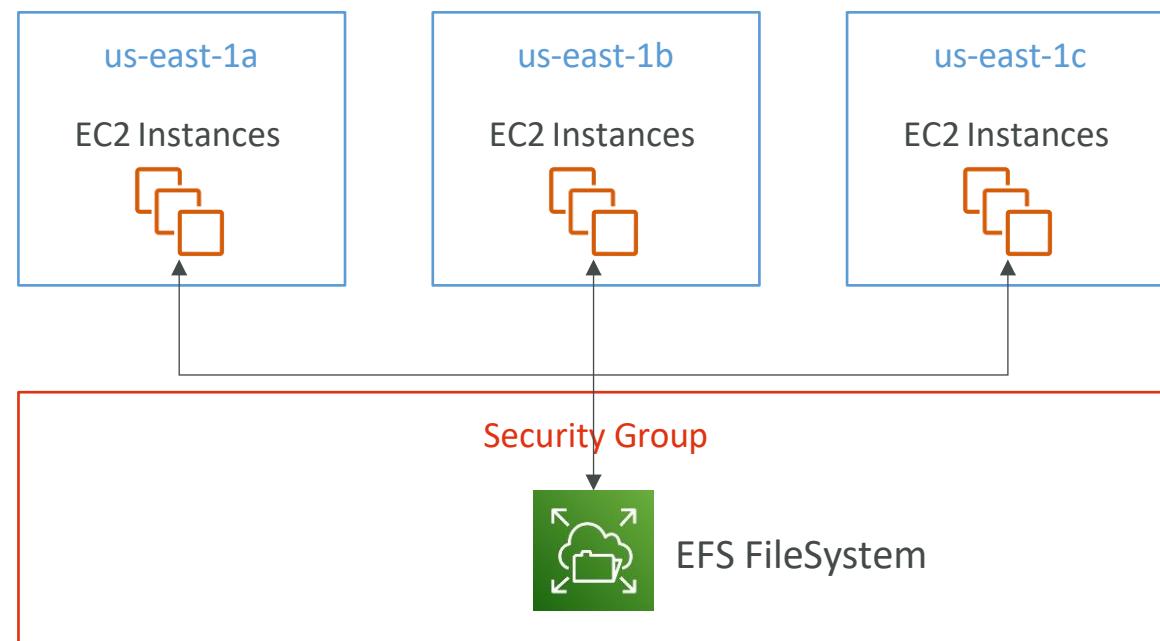
- When you create an encrypted EBS volume, you get the following:
 - Data at rest is encrypted inside the volume
 - All the data in flight moving between the instance and the volume is encrypted
 - All snapshots are encrypted
 - All volumes created from the snapshot
- Encryption and decryption are handled transparently (you have nothing to do)
- Encryption has a minimal impact on latency
- EBS Encryption leverages keys from KMS (AES-256)
- Copying an unencrypted snapshot allows encryption
- Snapshots of encrypted volumes are encrypted

Encryption: encrypt an unencrypted EBS volume

- Create an EBS snapshot of the volume
- Encrypt the EBS snapshot (using copy)
- Create new ebs volume from the snapshot (the volume will also be encrypted)
- Now you can attach the encrypted volume to the original instance

Amazon EFS - Elastic File System

- Managed NFS (network file system) that can be mounted on many EC2
- EFS works with EC2 instances in multi-AZ
- Highly available, scalable, expensive (3x gp2), pay per use



Amazon EFS - Elastic File System

- Use cases: content management, web serving, data sharing, Wordpress
- Uses NFSv4.1 protocol
- Uses security group to control access to EFS
- Compatible with Linux based AMI (not Windows)
- Encryption at rest using KMS

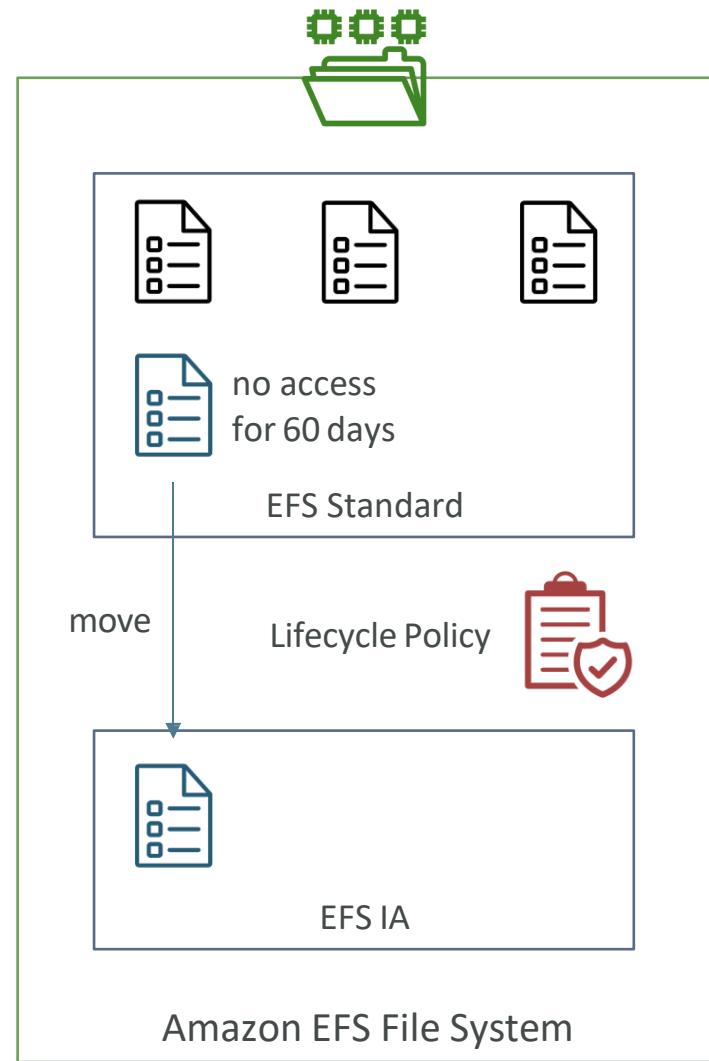
- POSIX file system (~Linux) that has a standard file API
- File system scales automatically, pay-per-use, no capacity planning!

EFS - Performance & Storage Classes

- EFS Scale
 - 1000s of concurrent NFS clients, 10 GB+ /s throughput
 - Grow to Petabyte-scale network file system, automatically
- Performance Mode (set at EFS creation time)
 - General Purpose (default) - latency-sensitive use cases (web server, CMS, etc...)
 - Max I/O - higher latency, throughput, highly parallel (big data, media processing)
- Throughput Mode
 - Bursting - 1 TB = 50MiB/s + burst of up to 100MiB/s
 - Provisioned - set your throughput regardless of storage size, ex: 1 GiB/s for 1 TB storage
 - Elastic - automatically scales throughput up or down based on your workloads
 - Up to 3GiB/s for reads and 1GiB/s for writes
 - Used for unpredictable workloads

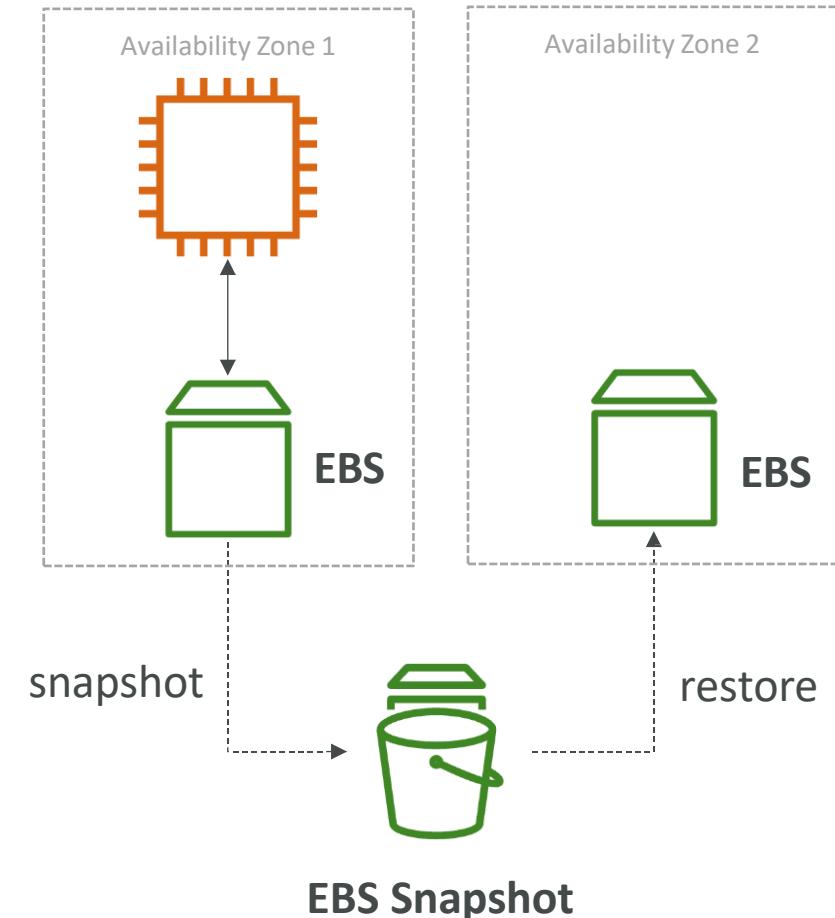
EFS - Storage Classes

- Storage Tiers (lifecycle management feature - move file after N days)
 - Standard: for frequently accessed files
 - Infrequent access (EFS-IA): cost to retrieve files, lower price to store.
 - Archive: rarely accessed data (few times each year), 50% cheaper
 - Implement lifecycle policies to move files between storage tiers
 - Availability and durability
 - Standard: Multi-AZ, great for prod
 - One Zone: One AZ, great for dev, backup enabled by default, compatible with IA (EFS One Zone-IA)
 - Over 90% in cost savings



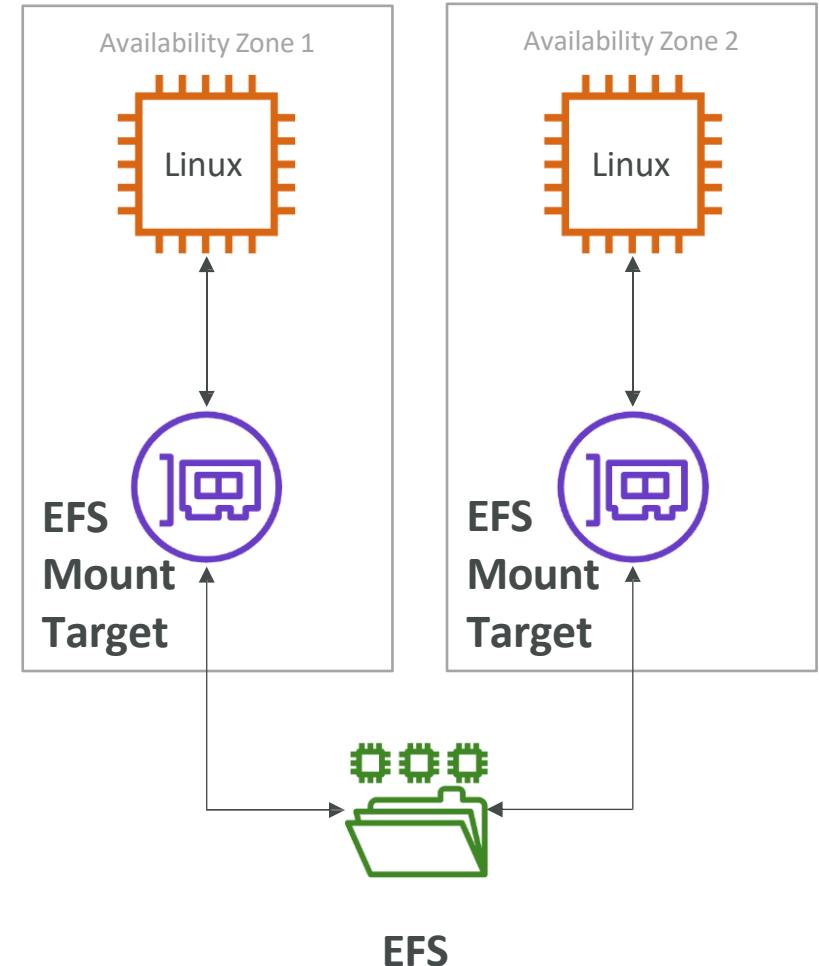
EBS vs EFS - Elastic Block Storage

- EBS volumes...
 - one instance (except multi-attach io1/io2)
 - are locked at the Availability Zone (AZ) level
 - gp2: IO increases if the disk size increases
 - gp3 & io1: can increase IO independently
- To migrate an EBS volume across AZ
 - Take a snapshot
 - Restore the snapshot to another AZ
 - EBS backups use IO and you shouldn't run them while your application is handling a lot of traffic
- Root EBS Volumes of instances get terminated by default if the EC2 instance gets terminated. (you can disable that)



EBS vs EFS - Elastic File System

- Mounting 100s of instances across AZ
 - EFS share website files (WordPress)
 - Only for Linux Instances (POSIX)
-
- EFS has a higher price point than EBS
 - Can leverage Storage Tiers for cost savings
-
- Remember: EFS vs EBS vs Instance Store



High Availability & Scalability

Scalability & High Availability

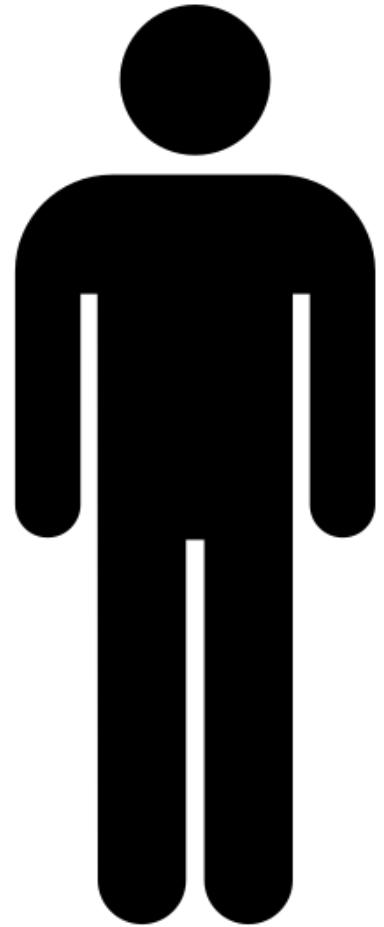
- Scalability means that an application / system can handle greater loads by adapting.
- There are two kinds of scalability:
 - Vertical Scalability
 - Horizontal Scalability (= elasticity)
- Scalability is linked but different to High Availability
- Let's deep dive into the distinction, using a call center as an example

Vertical Scalability

- Vertically scalability means increasing the size of the instance
- For example, your application runs on a t2.micro
- Scaling that application vertically means running it on a t2.large
- Vertical scalability is very common for non distributed systems, such as a database.
- RDS, ElastiCache are services that can scale vertically.
- There's usually a limit to how much you can vertically scale (hardware limit)



junior operator

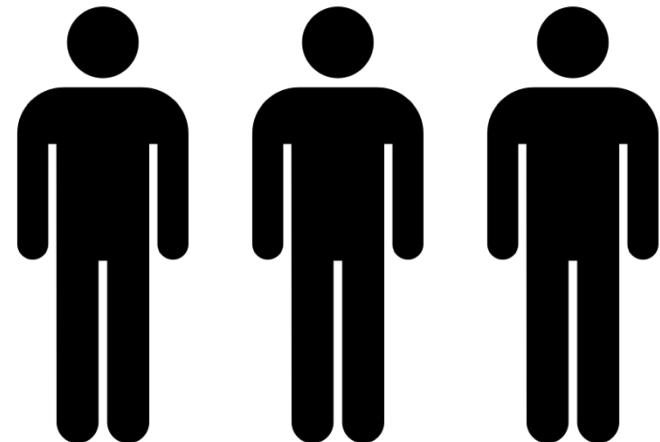
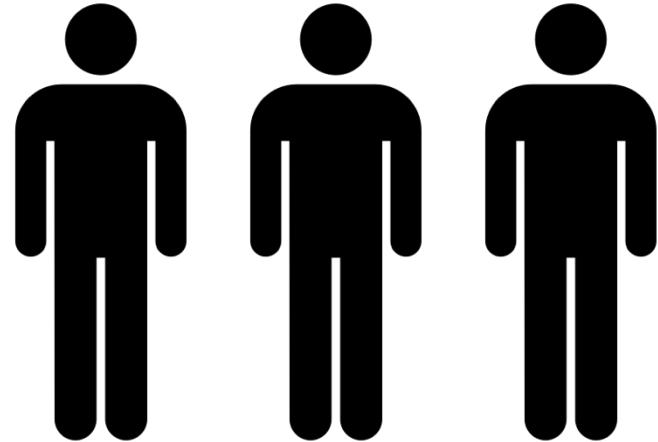


senior operator

Horizontal Scalability

- Horizontal Scalability means increasing the number of instances / systems for your application
- Horizontal scaling implies distributed systems.
- This is very common for web applications / modern applications
- It's easy to horizontally scale thanks the cloud offerings such as Amazon EC2

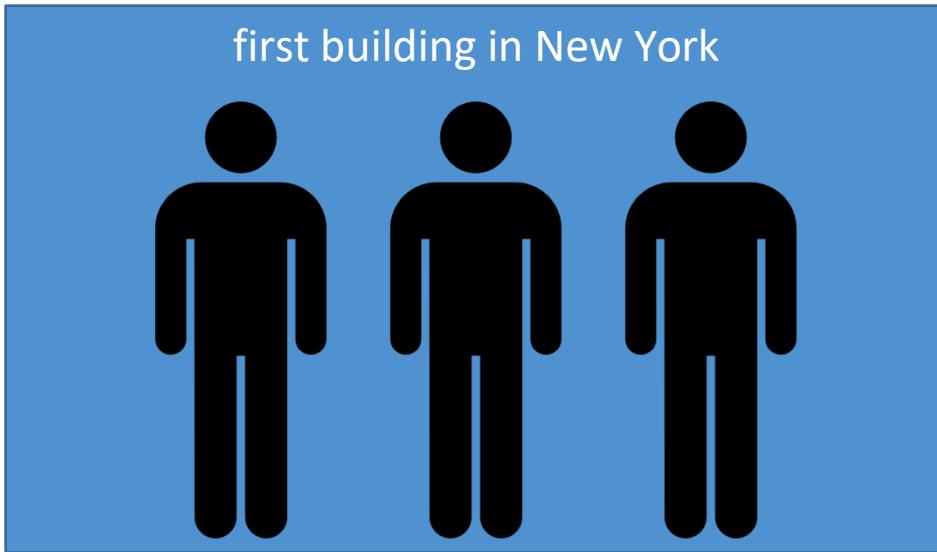
operator operator operator



operator operator operator

High Availability

- High Availability usually goes hand in hand with horizontal scaling
- High availability means running your application / system in at least 2 data centers (== Availability Zones)
- The goal of high availability is to survive a data center loss
- The high availability can be passive (for RDS Multi AZ for example)
- The high availability can be active (for horizontal scaling)

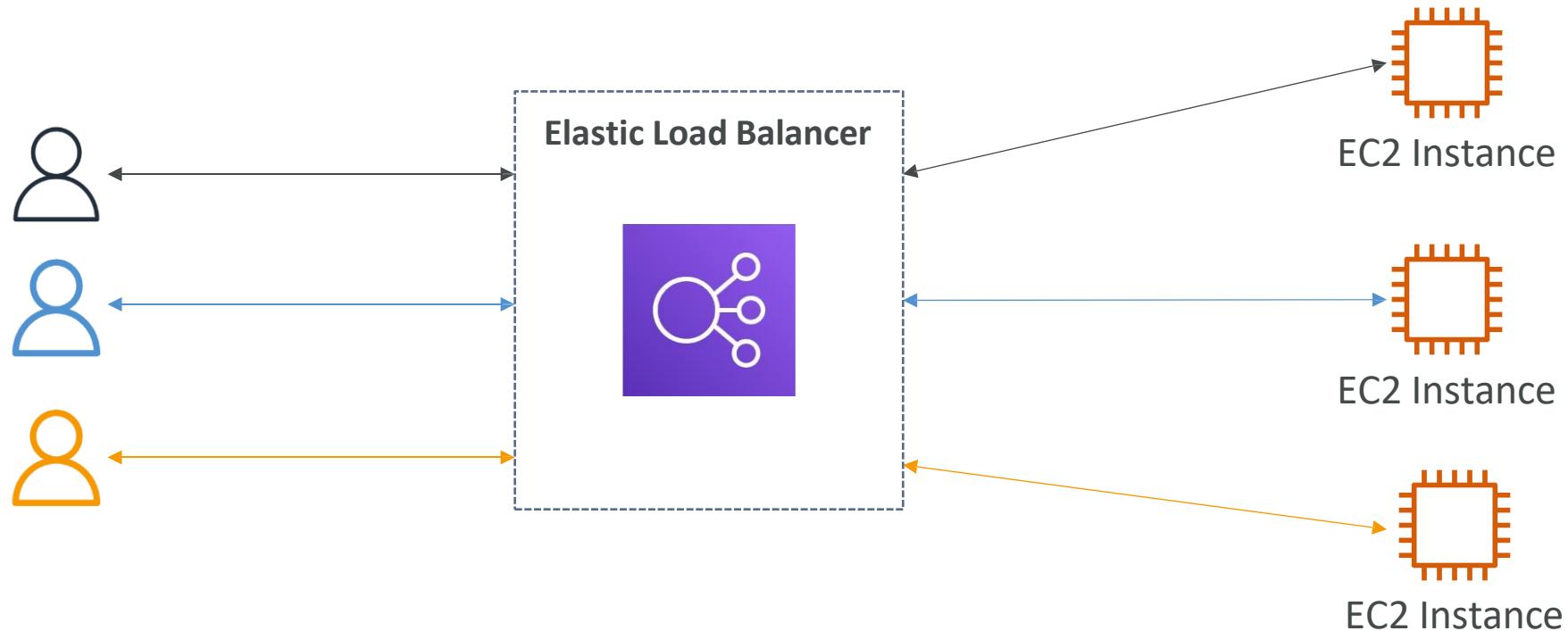


High Availability & Scalability For EC2

- Vertical Scaling: Increase instance size (= scale up / down)
 - From: t2.nano - 0.5G of RAM, 1 vCPU
 - To: u-12tb1.metal - 12.3 TB of RAM, 448 vCPUs
- Horizontal Scaling: Increase number of instances (= scale out / in)
 - Auto Scaling Group
 - Load Balancer
- High Availability: Run instances for the same application across multi AZ
 - Auto Scaling Group multi AZ
 - Load Balancer multi AZ

What is load balancing?

- Load Balancers are servers that forward traffic to multiple servers (e.g., EC2 instances) downstream



Why use a load balancer?

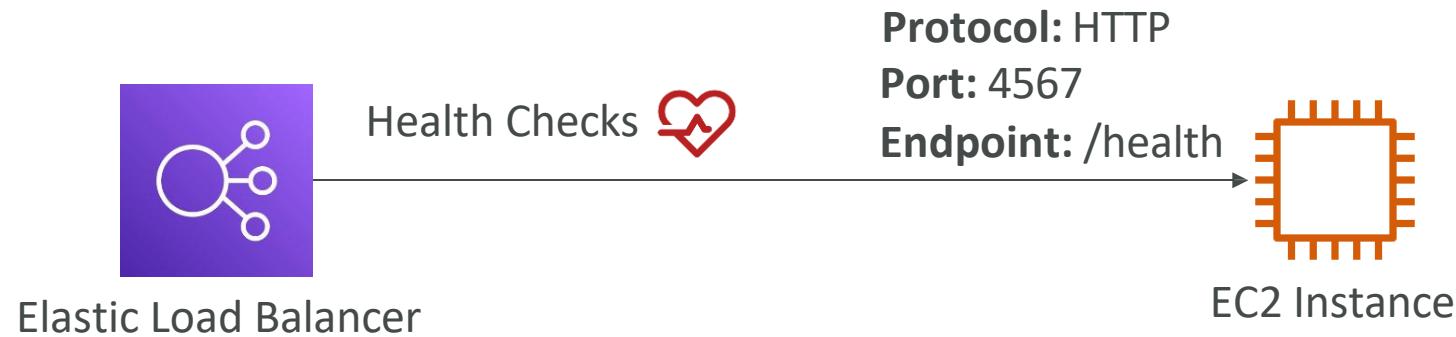
- Spread load across multiple downstream instances
- Expose a single point of access (DNS) to your application
- Seamlessly handle failures of downstream instances
- Do regular health checks to your instances
- Provide SSL termination (HTTPS) for your websites
- Enforce stickiness with cookies
- High availability across zones
- Separate public traffic from private traffic

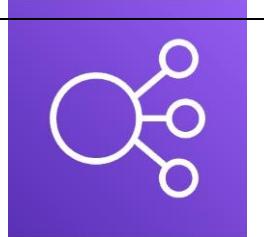
Why use an Elastic Load Balancer?

- An Elastic Load Balancer is a [managed load balancer](#)
 - AWS guarantees that it will be working
 - AWS takes care of upgrades, maintenance, high availability
 - AWS provides only a few configuration knobs
- It costs less to setup your own load balancer but it will be a lot more effort on your end
- It is integrated with many AWS offerings / services
 - EC2, EC2 Auto Scaling Groups, Amazon ECS
 - AWS Certificate Manager (ACM), CloudWatch
 - Route 53, AWS WAF, AWS Global Accelerator

Health Checks

- Health Checks are crucial for Load Balancers
- They enable the load balancer to know if instances it forwards traffic to are available to reply to requests
- The health check is done on a port and a route (`/health` is common)
- If the response is not 200 (OK), then the instance is unhealthy





Types of load balancer on AWS

- AWS has **4 kinds of managed Load Balancers**
- Classic Load Balancer (v1 - old generation) - 2009 - CLB
 - HTTP, HTTPS, TCP, SSL (secure TCP)
- Application Load Balancer (v2 - new generation) - 2016 - ALB
 - HTTP, HTTPS, WebSocket
- Network Load Balancer (v2 - new generation) - 2017 - NLB
 - TCP, TLS (secure TCP), UDP
- Gateway Load Balancer - 2020 - GWLB
 - Operates at layer 3 (Network layer) - IP Protocol
- Overall, it is recommended to use the newer generation load balancers as they provide more features
- Some load balancers can be setup as **internal** (private) or **external** (public) ELBs

Load Balancer Security Groups



Load Balancer Security Group:

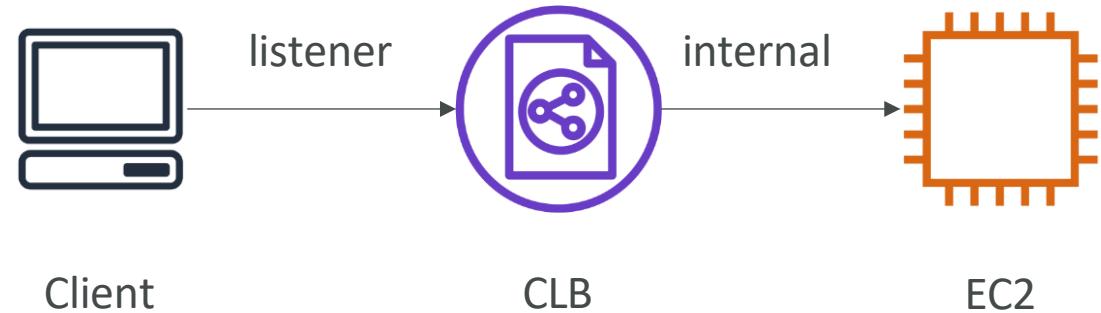
Type <small>i</small>	Protocol <small>i</small>	Port Range <small>i</small>	Source <small>i</small>	Description <small>i</small>
HTTP	TCP	80	0.0.0.0/0	Allow HTTP from an...
HTTPS	TCP	443	0.0.0.0/0	Allow HTTPS from a...

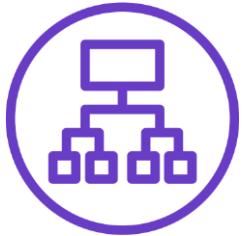
Application Security Group: Allow traffic only from Load Balancer

Type <small>i</small>	Protocol <small>i</small>	Port Range <small>i</small>	Source <small>i</small>	Description <small>i</small>
HTTP	TCP	80	sg-054b5ff5ea02f2b6e (load-b	Allow Traffic only...

Classic Load Balancers (v1)

- Supports TCP (Layer 4), HTTP & HTTPS (Layer 7)
- Health checks are TCP or HTTP based
- Fixed hostname
XXX.region.elb.amazonaws.com

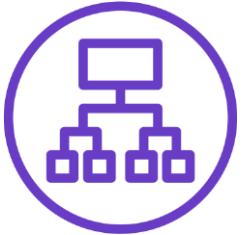




Application Load Balancer (v2)

- Application load balancers is Layer 7 (HTTP)
- Load balancing to multiple HTTP applications across machines (target groups)
- Load balancing to multiple applications on the same machine (ex: containers)
- Support for HTTP/2 and WebSocket
- Support redirects (from HTTP to HTTPS for example)

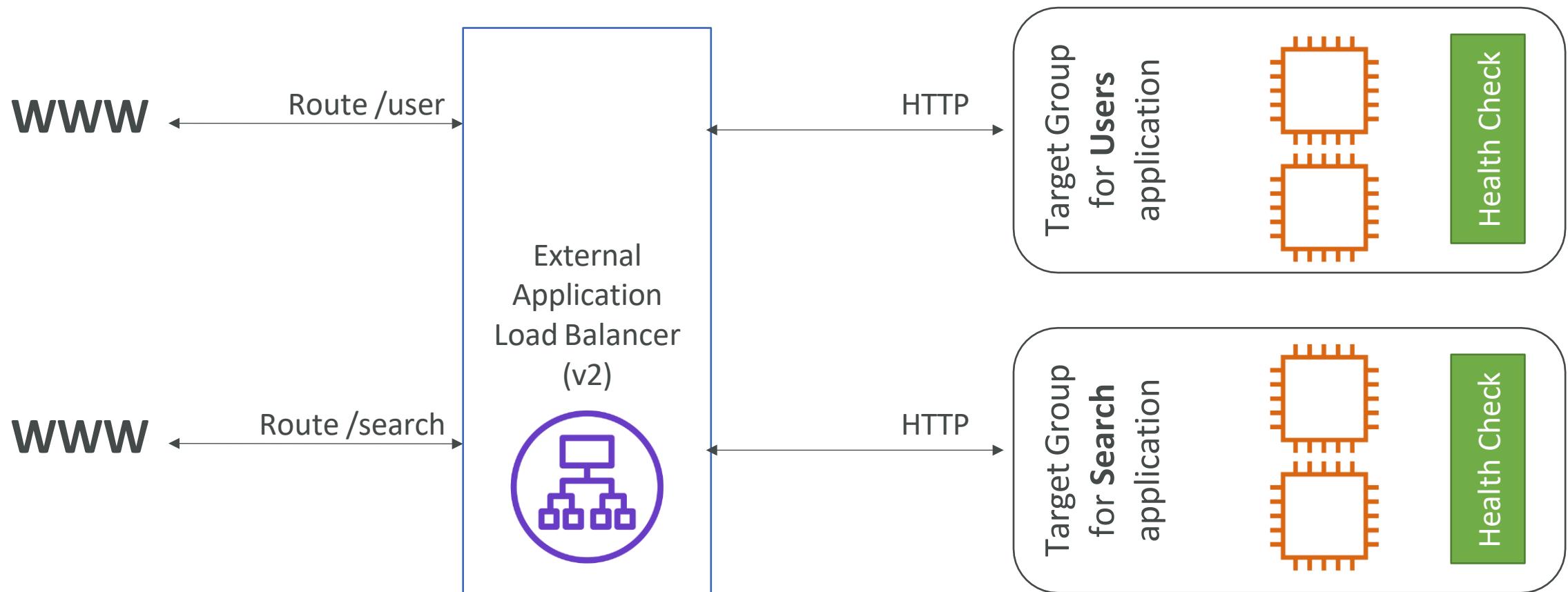
Application Load Balancer (v2)



- Routing tables to different target groups:
 - Routing based on path in URL (example.com/users & example.com/posts)
 - Routing based on hostname in URL (one.example.com & other.example.com)
 - Routing based on Query String, Headers
(example.com/users?id=123&order=false)
- ALB are a great fit for micro services & container-based application
(example: Docker & Amazon ECS)
- Has a port mapping feature to redirect to a dynamic port in ECS
- In comparison, we'd need multiple Classic Load Balancer per application

Application Load Balancer (v2)

HTTP Based Traffic



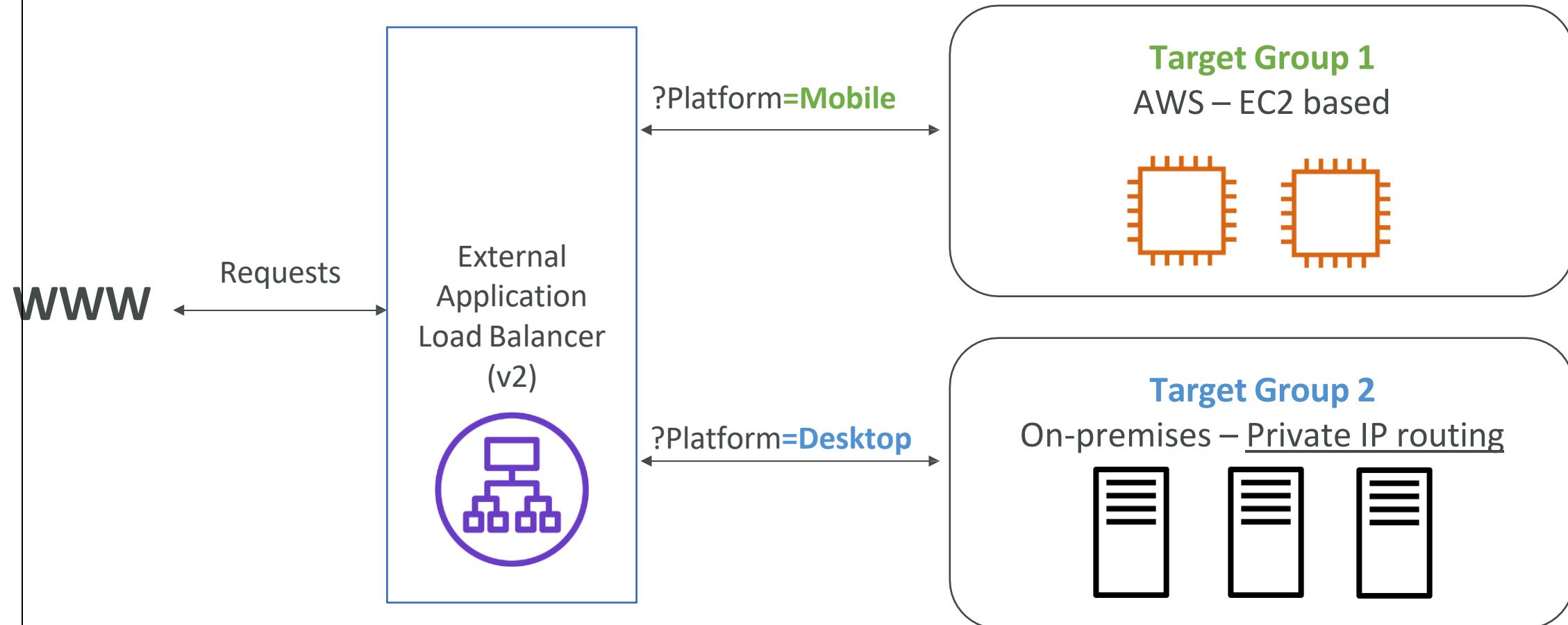
Application Load Balancer (v2)

Target Groups

- EC2 instances (can be managed by an Auto Scaling Group) - HTTP
 - ECS tasks (managed by ECS itself) - HTTP
 - Lambda functions - HTTP request is translated into a JSON event
 - IP Addresses - must be private IPs
-
- ALB can route to multiple target groups
 - Health checks are at the target group level

Application Load Balancer (v2)

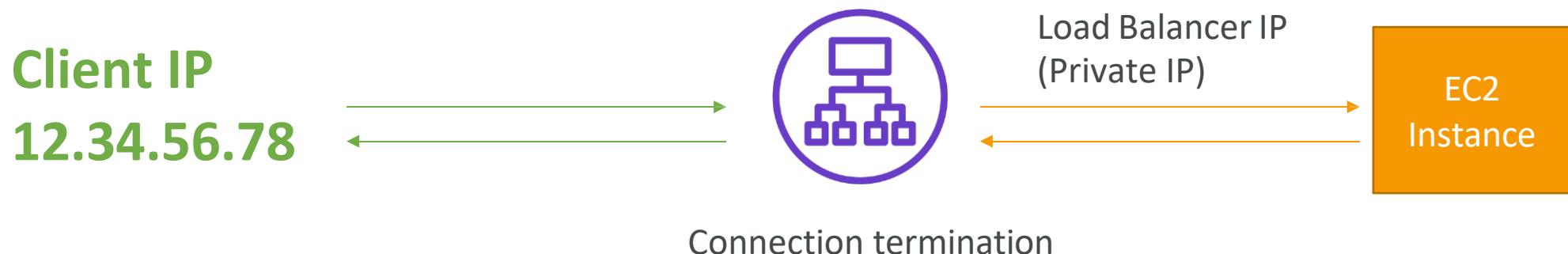
Query Strings/Parameters Routing

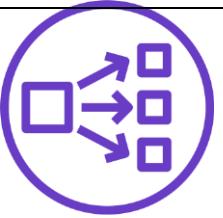


Application Load Balancer (v2)

Good to Know

- Fixed hostname (XXX.region.elb.amazonaws.com)
- The application servers don't see the IP of the client directly
 - The true IP of the client is inserted in the header X-Forwarded-For
 - We can also get Port (X-Forwarded-Port) and proto (X-Forwarded-Proto)



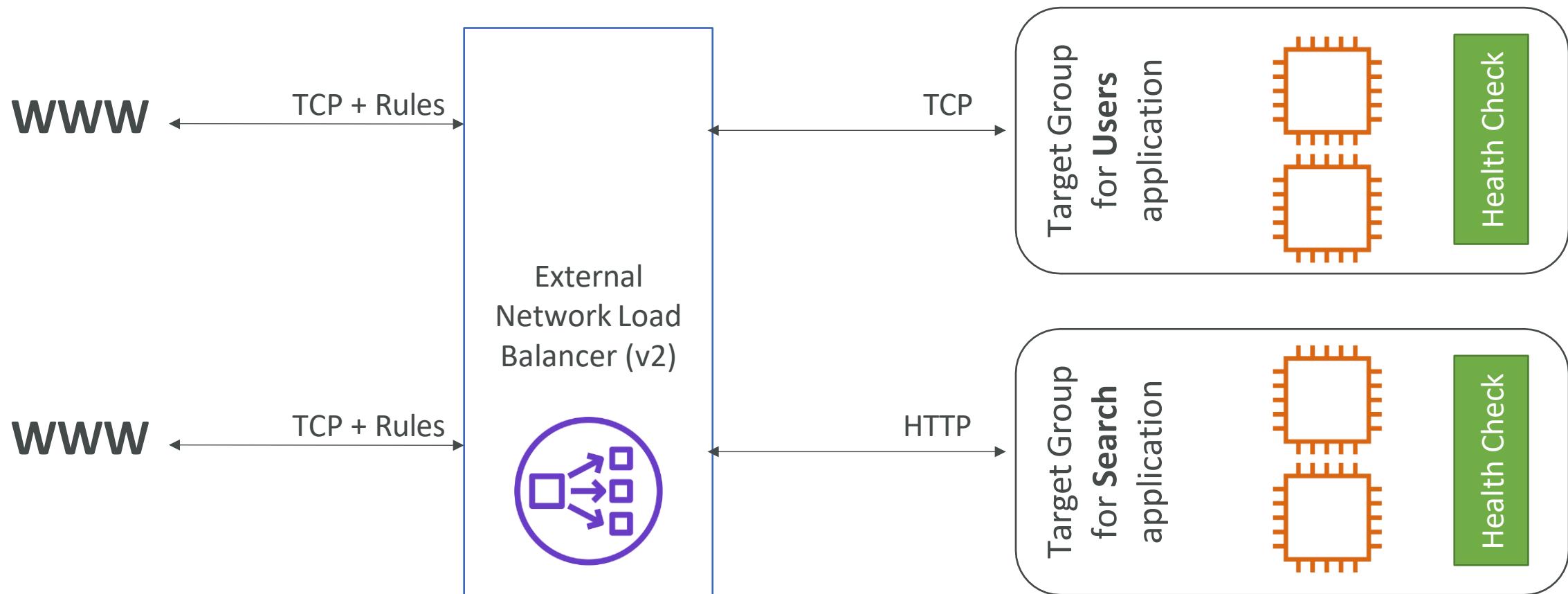


Network Load Balancer (v2)

- Network load balancers (Layer 4) allow to:
 - Forward TCP & UDP traffic to your instances
 - Handle millions of requests per second
 - Ultra-low latency
- NLB has one static IP per AZ, and supports assigning Elastic IP (helpful for whitelisting specific IP)
- NLB are used for extreme performance, TCP or UDP traffic

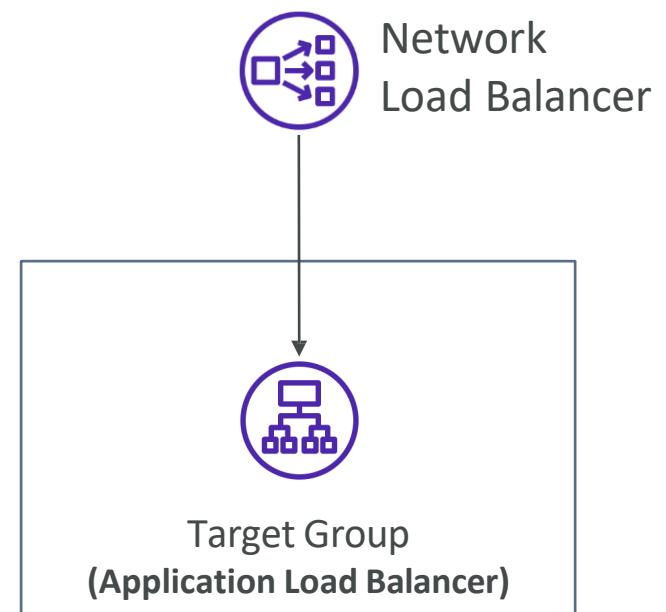
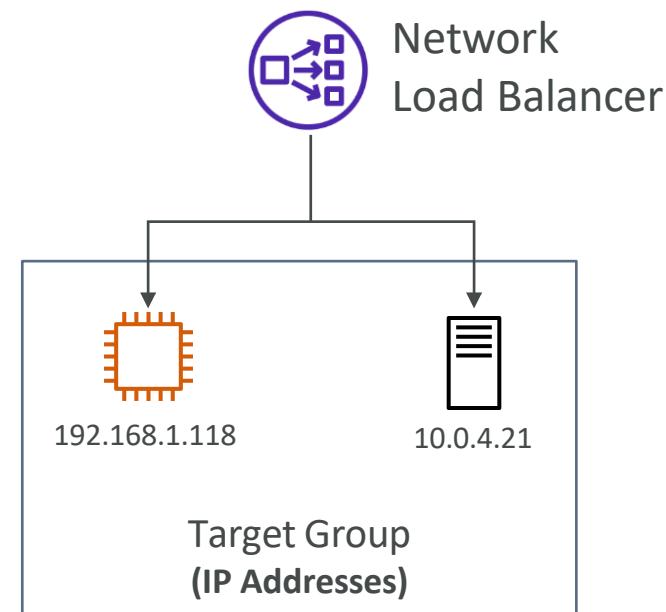
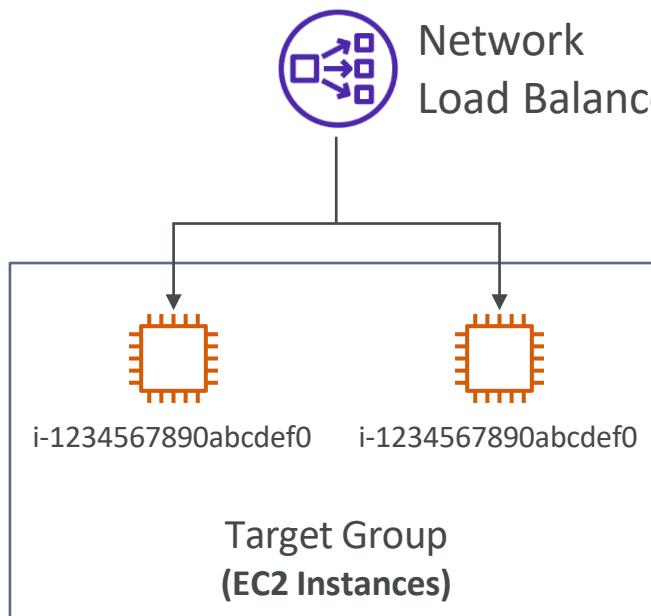
Network Load Balancer (v2)

TCP (Layer 4) Based Traffic



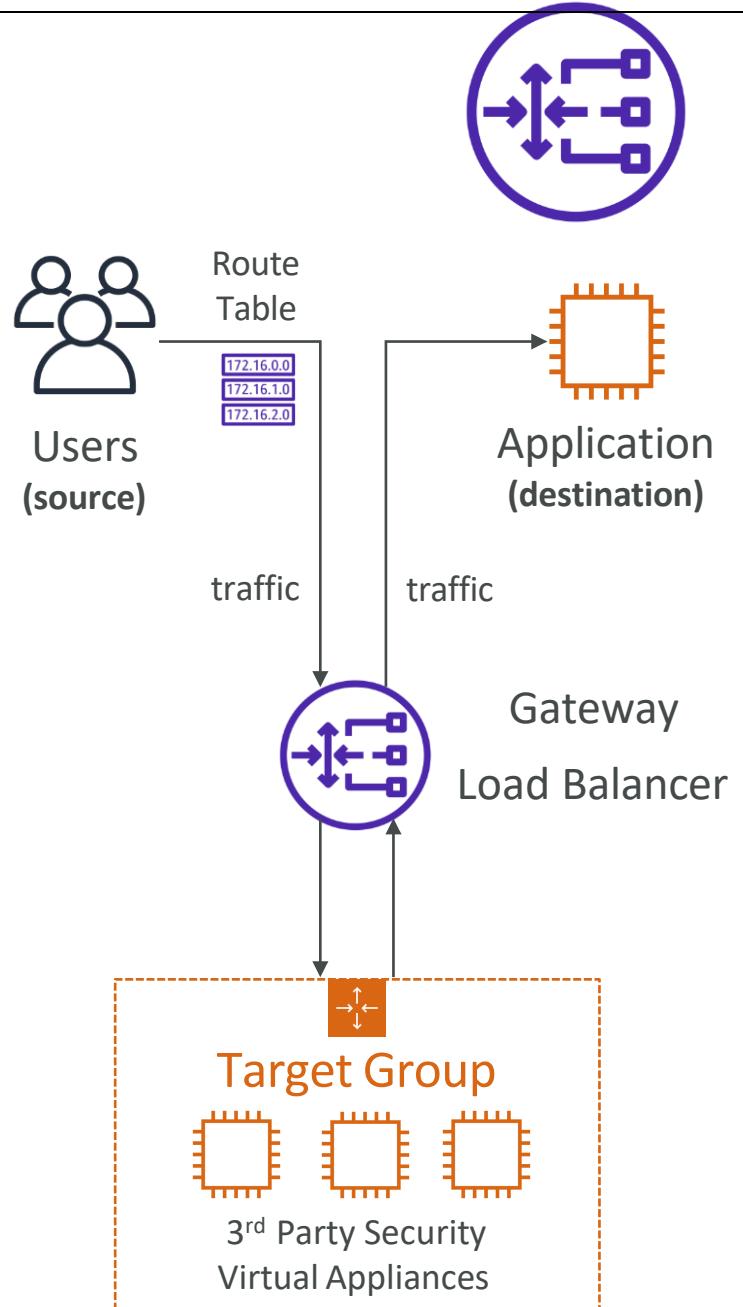
Network Load Balancer - Target Groups

- EC2 instances
- IP Addresses - must be private IPs
- Application Load Balancer
- Health Checks support the TCP, HTTP and HTTPS Protocols



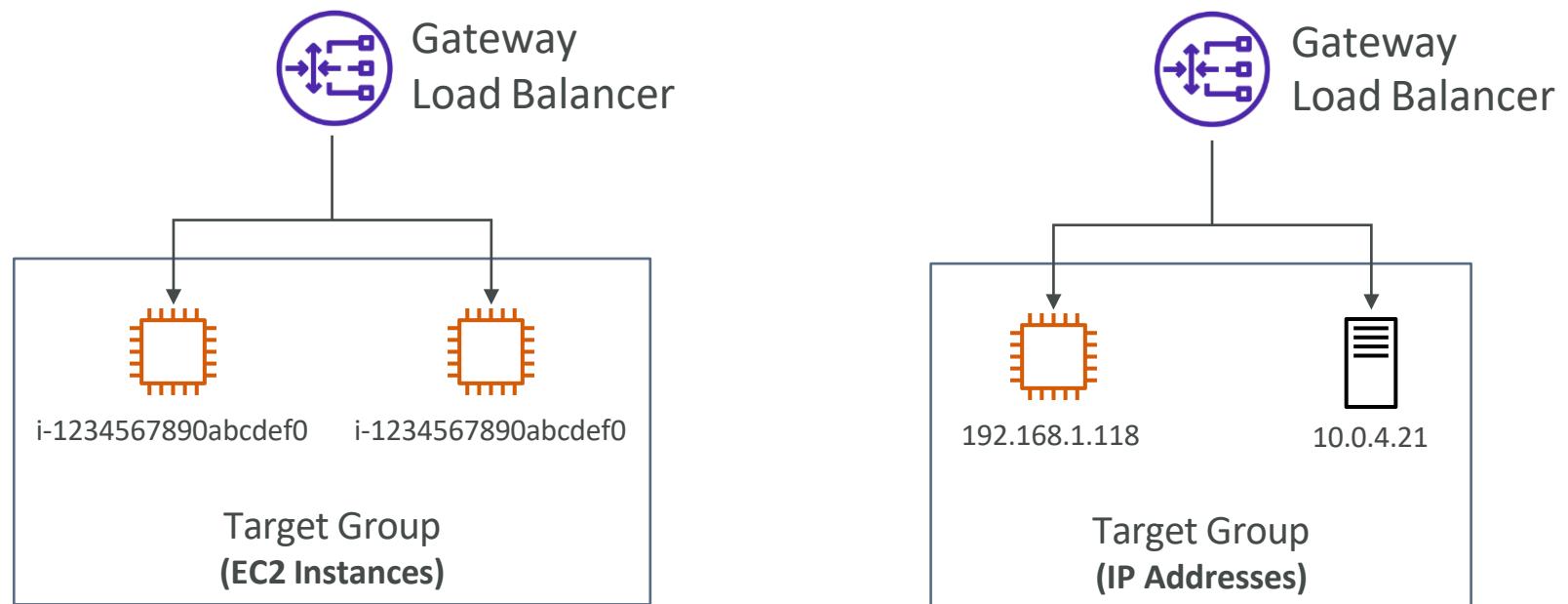
Gateway Load Balancer

- Deploy, scale, and manage a fleet of 3rd party network virtual appliances in AWS
- Example: Firewalls, Intrusion Detection and Prevention Systems, Deep Packet Inspection Systems, payload manipulation, ...
- Operates at Layer 3 (Network Layer) - IP Packets
- Combines the following functions:
 - Transparent Network Gateway - single entry/exit for all traffic
 - Load Balancer - distributes traffic to your virtual appliances
- Uses the GENEVE protocol on port 6081



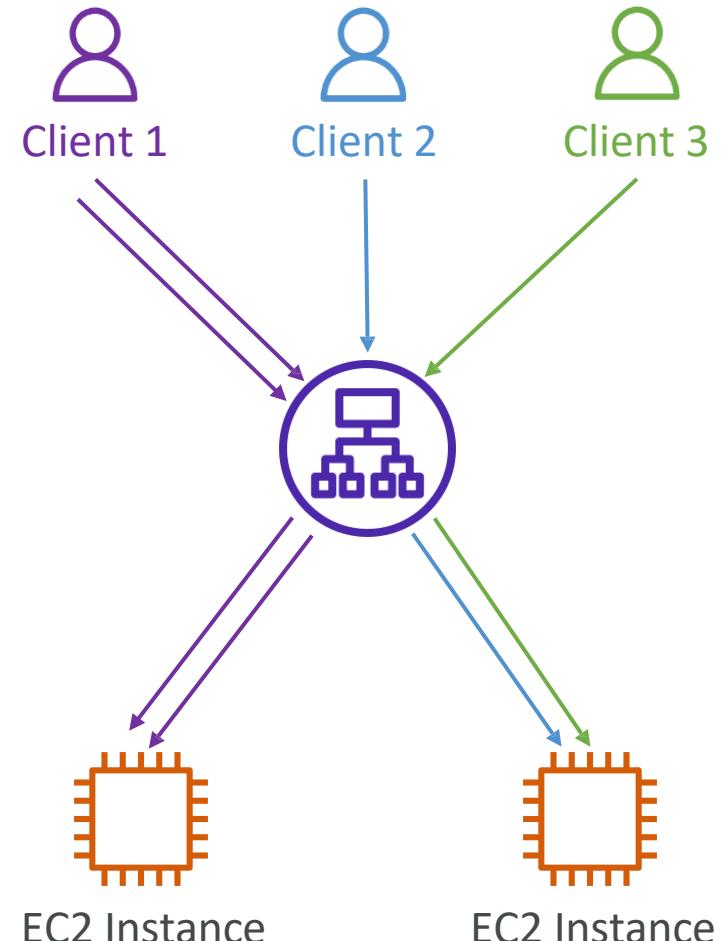
Gateway Load Balancer - Target Groups

- EC2 instances
- IP Addresses - must be private IPs



Sticky Sessions (Session Affinity)

- It is possible to implement stickiness so that the same client is always redirected to the same instance behind a load balancer
- This works for Classic Load Balancer, Application Load Balancer, and Network Load Balancer
- For both CLB & ALB, the “cookie” used for stickiness has an expiration date you control
- Use case: make sure the user doesn’t lose his session data
- Enabling stickiness may bring imbalance to the load over the backend EC2 instances



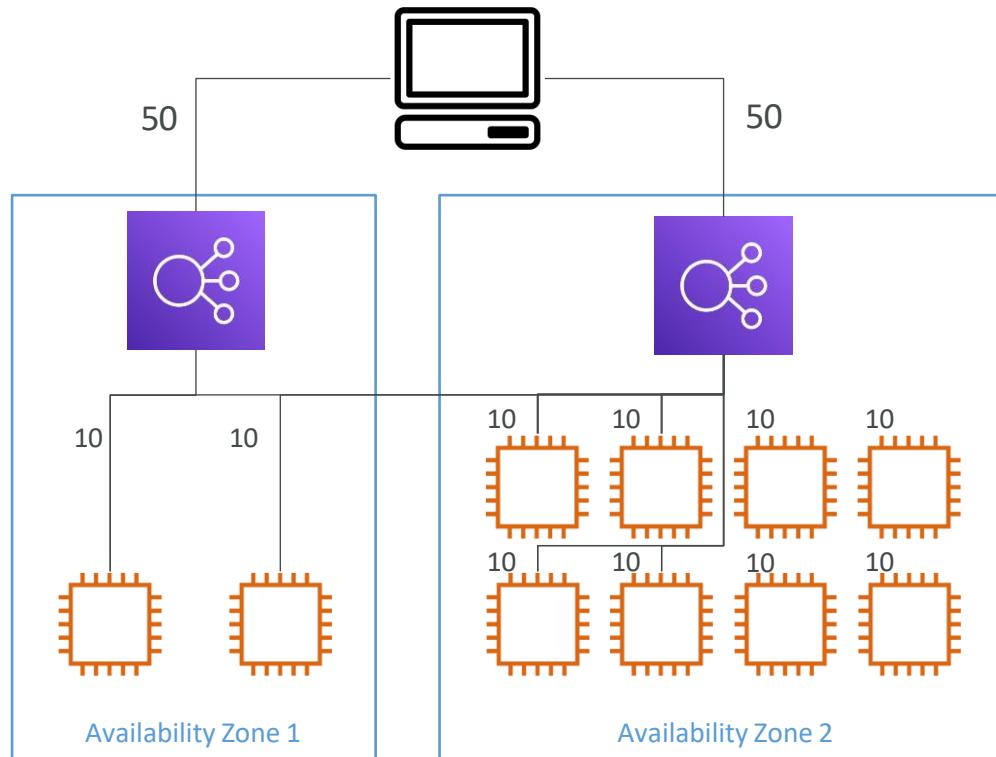
Sticky Sessions - Cookie Names

- Application-based Cookies
 - Custom cookie
 - Generated by the target
 - Can include any custom attributes required by the application
 - Cookie name must be specified individually for each target group
 - Don't use AWSALB, AWSALBAPP, or AWSALBTG (reserved for use by the ELB)
 - Application cookie
 - Generated by the load balancer
 - Cookie name is AWSALBAPP
- Duration-based Cookies
 - Cookie generated by the load balancer
 - Cookie name is AWSALB for ALB, AWSELB for CLB

Cross-Zone Load Balancing

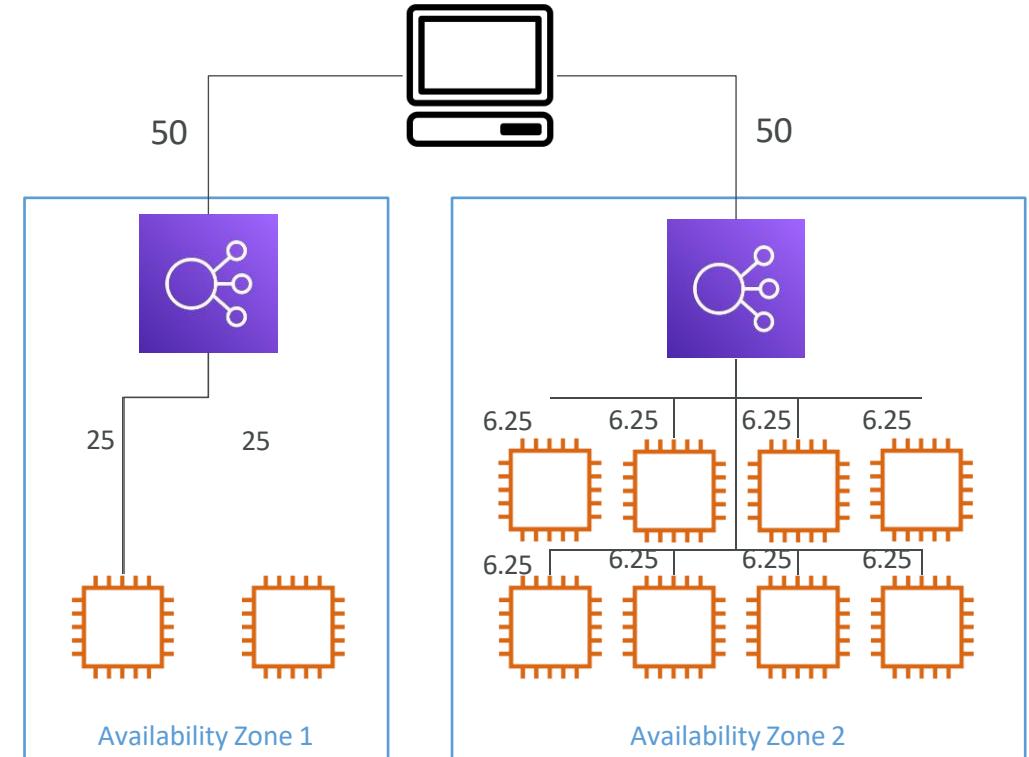
With Cross Zone Load Balancing:

each load balancer instance distributes evenly across all registered instances in all AZ



Without Cross Zone Load Balancing:

Requests are distributed in the instances of the node of the Elastic Load Balancer



Cross-Zone Load Balancing

- Application Load Balancer
 - Enabled by default (can be disabled at the Target Group level)
 - No charges for inter AZ data
- Network Load Balancer & Gateway Load Balancer
 - Disabled by default
 - You pay charges (\$) for inter AZ data if enabled
- Classic Load Balancer
 - Disabled by default
 - No charges for inter AZ data if enabled

SSL/TLS - Basics

- An SSL Certificate allows traffic between your clients and your load balancer to be encrypted in transit (in-flight encryption)
- SSL refers to Secure Sockets Layer, used to encrypt connections
- TLS refers to Transport Layer Security, which is a newer version
- Nowadays, TLS certificates are mainly used, but people still refer as SSL
- Public SSL certificates are issued by Certificate Authorities (CA)
- Comodo, Symantec, GoDaddy, GlobalSign, DigiCert, LetsEncrypt, etc...
- SSL certificates have an expiration date (you set) and must be renewed

Load Balancer - SSL Certificates



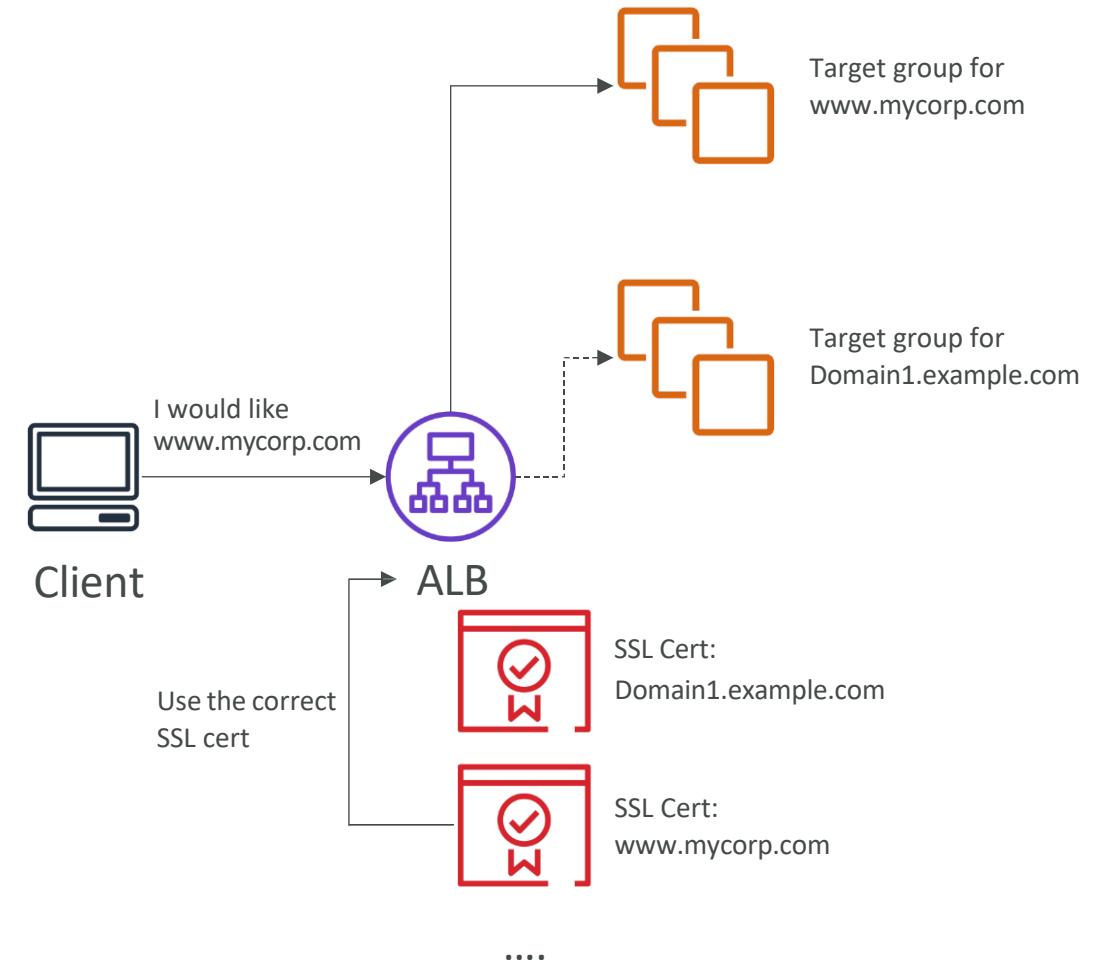
- The load balancer uses an X.509 certificate (SSL/TLS server certificate)
- You can manage certificates using ACM (AWS Certificate Manager)
- You can create/upload your own certificates alternatively
- **HTTPS listener:**
 - You must specify a default certificate
 - You can add an optional list of certs to support multiple domains
 - Clients can use SNI (Server Name Indication) to specify the hostname they reach
 - Ability to specify a security policy to support older versions of SSL / TLS (legacy clients)

SSL - Server Name Indication (SNI)

- SNI solves the problem of loading multiple SSL certificates onto one web server (to serve multiple websites)
- It's a "newer" protocol, and requires the client to indicate the hostname of the target server in the initial SSL handshake
- The server will then find the correct certificate, or return the default one

Note:

- Only works for ALB & NLB (newer generation), CloudFront
- Does not work for CLB (older gen)

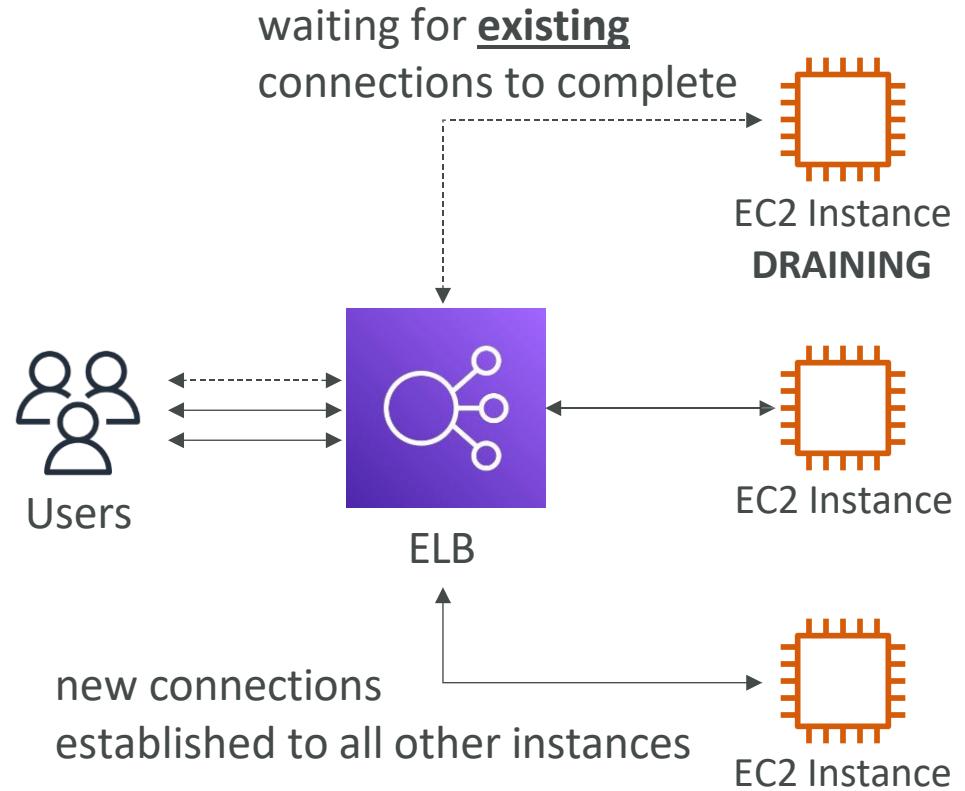


Elastic Load Balancers - SSL Certificates

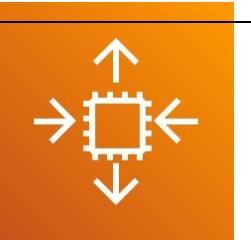
- Classic Load Balancer (v1)
 - Support only one SSL certificate
 - Must use multiple CLB for multiple hostname with multiple SSL certificates
- Application Load Balancer (v2)
 - Supports multiple listeners with multiple SSL certificates
 - Uses Server Name Indication (SNI) to make it work
- Network Load Balancer (v2)
 - Supports multiple listeners with multiple SSL certificates
 - Uses Server Name Indication (SNI) to make it work

Connection Draining

- Feature naming
 - Connection Draining - for CLB
 - Deregistration Delay - for ALB & NLB
- Time to complete “in-flight requests” while the instance is de-registering or unhealthy
- Stops sending new requests to the EC2 instance which is de-registering
- Between 1 to 3600 seconds (default: 300 seconds)
- Can be disabled (set value to 0)
- Set to a low value if your requests are short

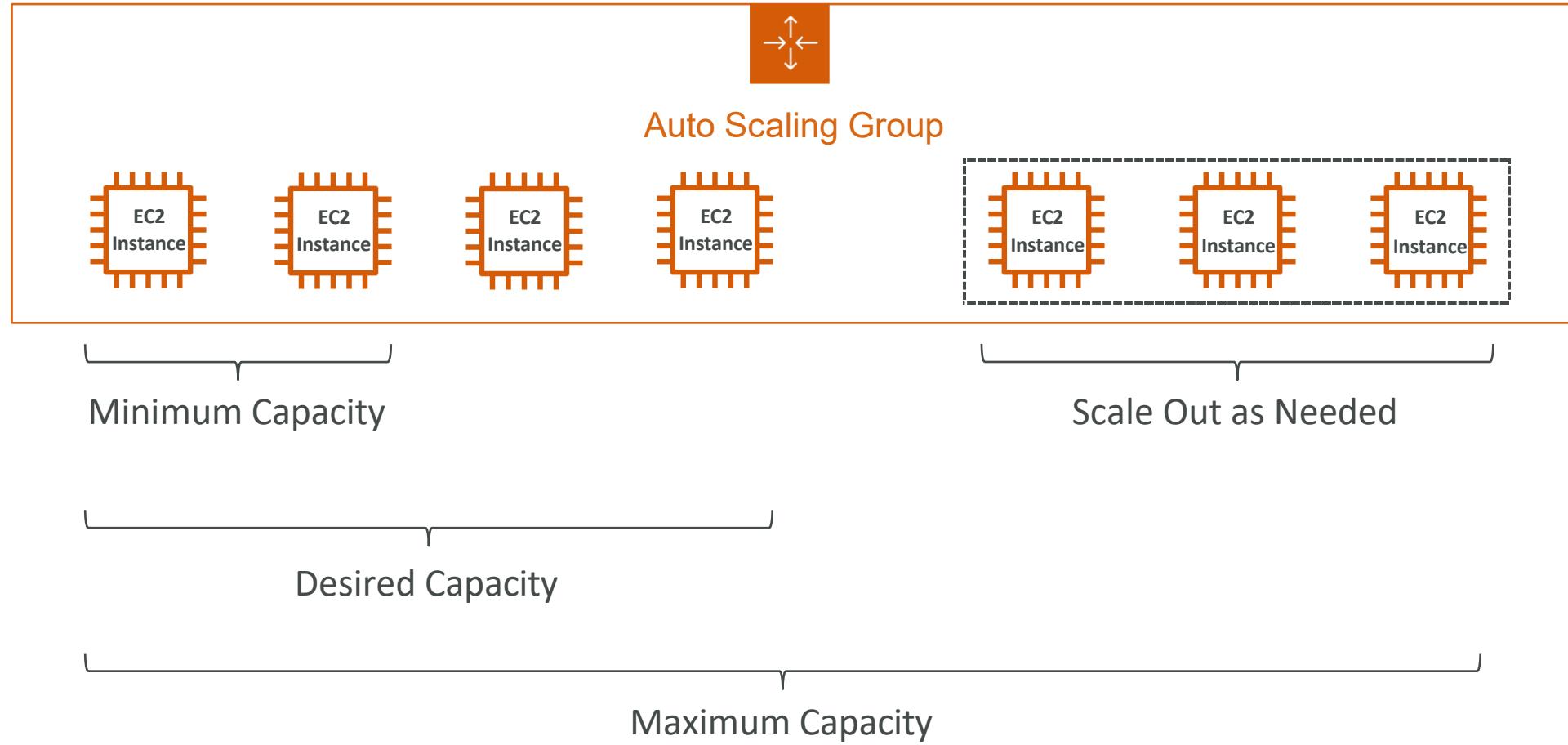


What's an Auto Scaling Group?

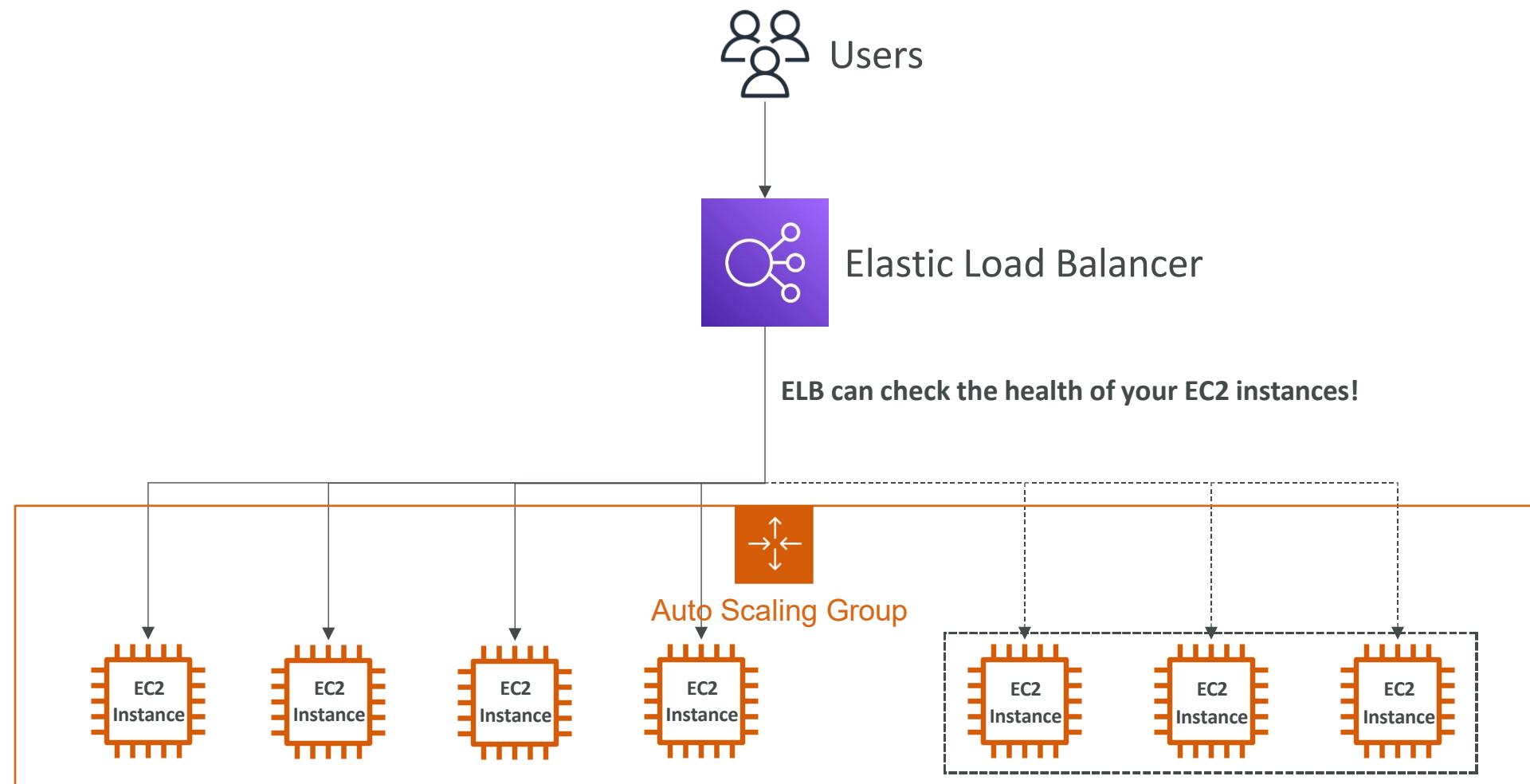


- In real-life, the load on your websites and application can change
- In the cloud, you can create and get rid of servers very quickly
- The goal of an Auto Scaling Group (ASG) is to:
 - Scale out (add EC2 instances) to match an increased load
 - Scale in (remove EC2 instances) to match a decreased load
 - Ensure we have a minimum and a maximum number of EC2 instances running
 - Automatically register new instances to a load balancer
 - Re-create an EC2 instance in case a previous one is terminated (ex: if unhealthy)
- ASG are free (you only pay for the underlying EC2 instances)

Auto Scaling Group in AWS

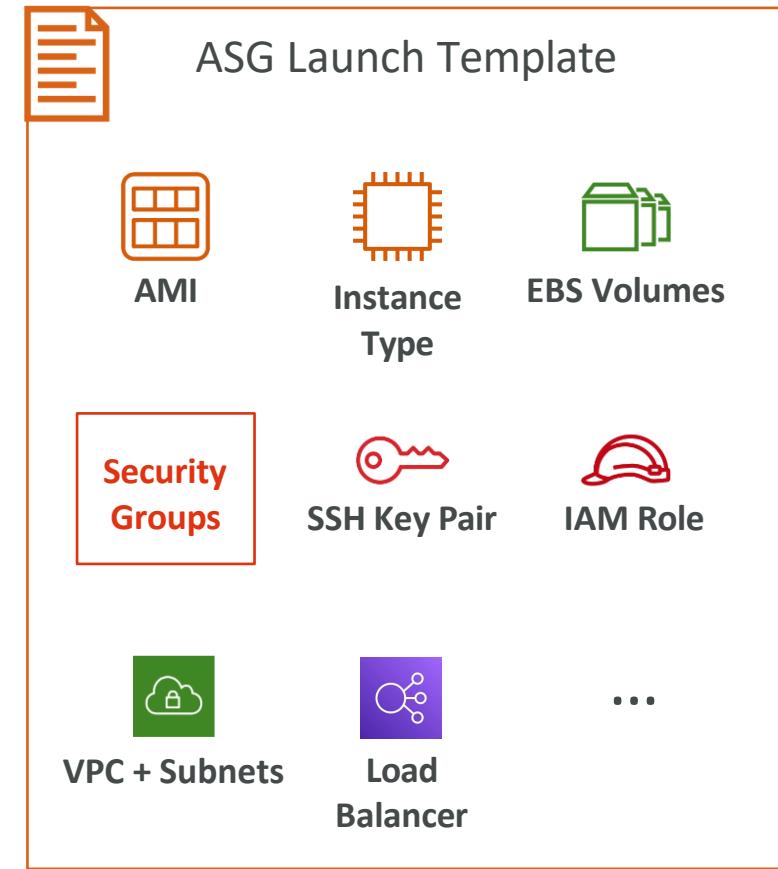


Auto Scaling Group in AWS With Load Balancer



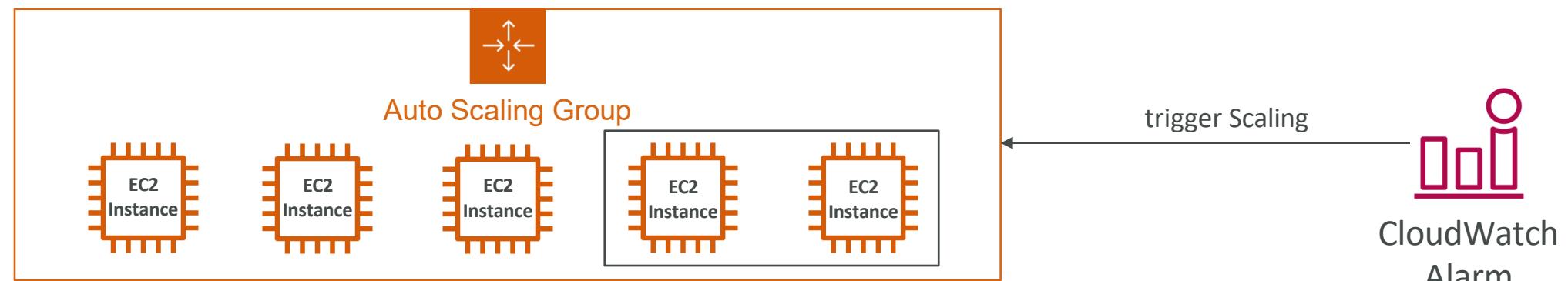
Auto Scaling Group Attributes

- A Launch Template (older “Launch Configurations” are deprecated)
 - AMI + Instance Type
 - EC2 User Data
 - EBS Volumes
 - Security Groups
 - SSH Key Pair
 - IAM Roles for your EC2 Instances
 - Network + Subnets Information
 - Load Balancer Information
- Min Size / Max Size / Initial Capacity
- Scaling Policies



Auto Scaling - CloudWatch Alarms & Scaling

- It is possible to scale an ASG based on CloudWatch alarms
- An alarm monitors a metric (such as Average CPU, or a custom metric)
- Metrics such as Average CPU are computed for the overall ASG instances
- Based on the alarm:
 - We can create scale-out policies (increase the number of instances)
 - We can create scale-in policies (decrease the number of instances)

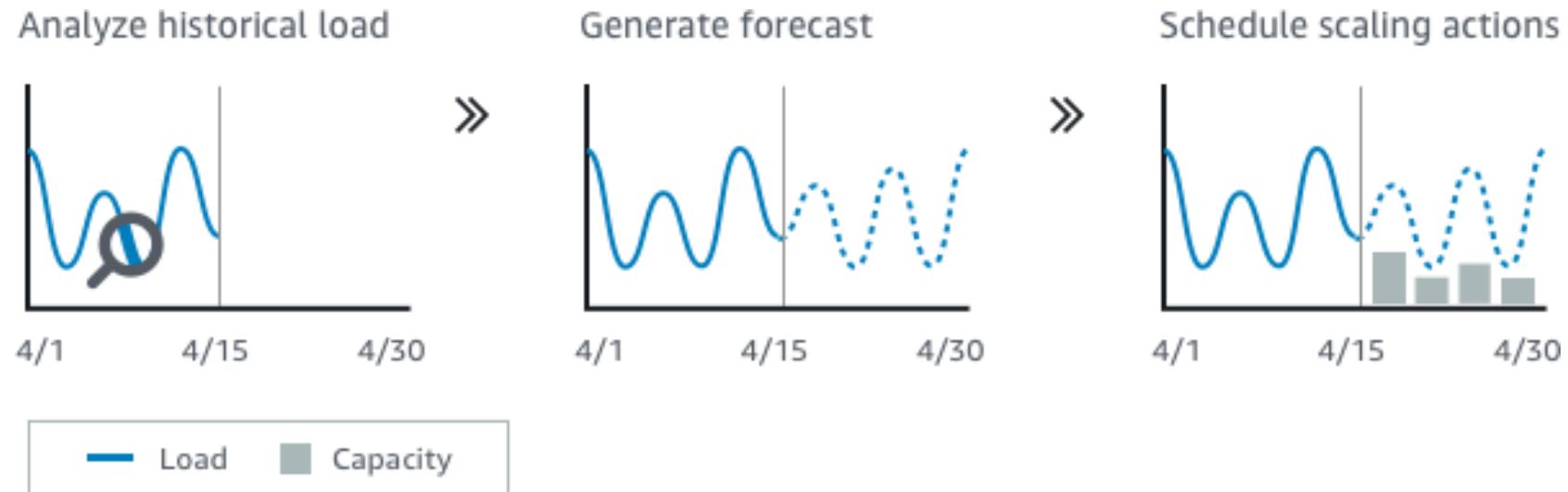


Auto Scaling Groups - Scaling Policies

- Dynamic Scaling
 - Target Tracking Scaling
 - Simple to set-up
 - Example: I want the average ASG CPU to stay at around 40%
 - Simple / Step Scaling
 - When a CloudWatch alarm is triggered (example CPU > 70%), then add 2 units
 - When a CloudWatch alarm is triggered (example CPU < 30%), then remove 1
- Scheduled Scaling
 - Anticipate a scaling based on known usage patterns
 - Example: increase the min capacity to 10 at 5 pm on Fridays

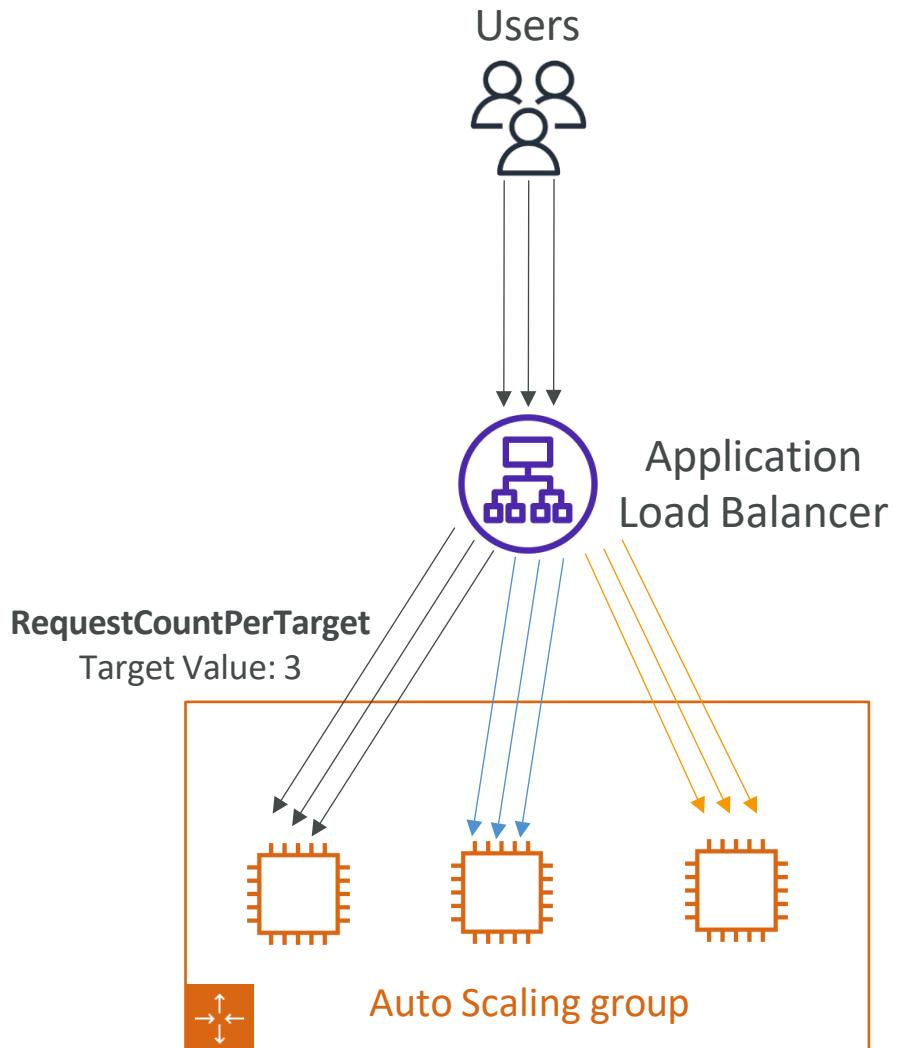
Auto Scaling Groups - Scaling Policies

- Predictive scaling: continuously forecast load and schedule scaling ahead



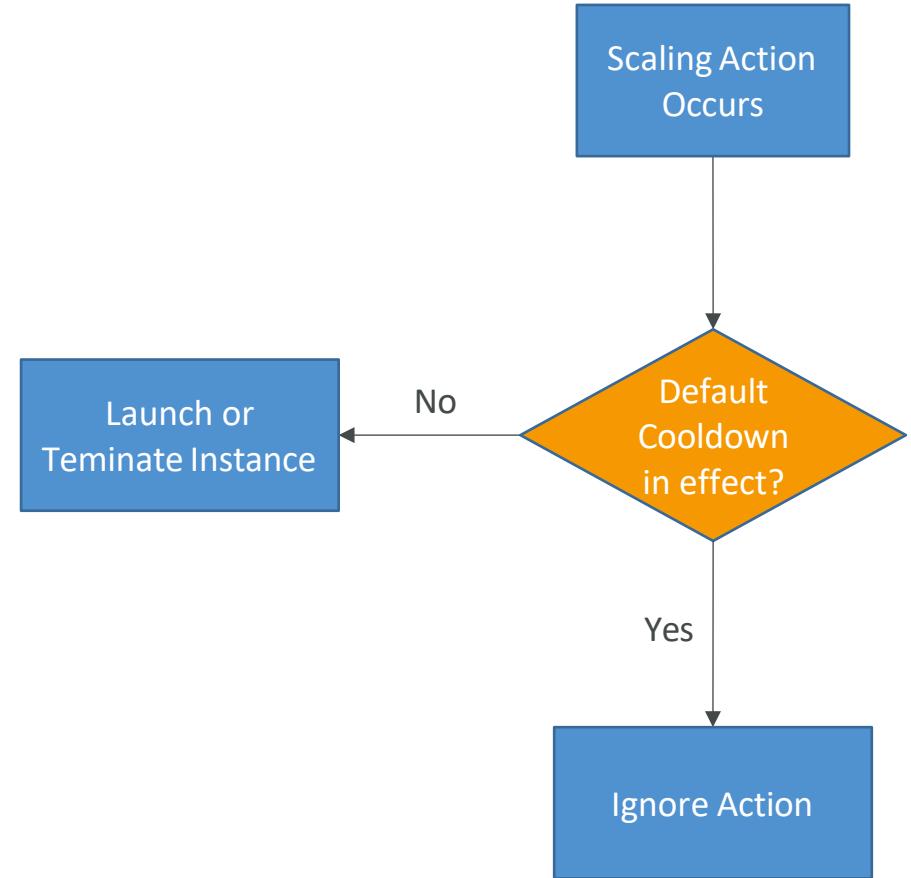
Good metrics to scale on

- CPUUtilization: Average CPU utilization across your instances
- RequestCountPerTarget: to make sure the number of requests per EC2 instances is stable
- Average Network In / Out (if you're application is network bound)
- Any custom metric (that you push using CloudWatch)



Auto Scaling Groups - Scaling Cooldowns

- After a scaling activity happens, you are in the cooldown period (default 300 seconds)
- During the cooldown period, the ASG will not launch or terminate additional instances (to allow for metrics to stabilize)
- Advice: Use a ready-to-use AMI to reduce configuration time in order to be serving request faster and reduce the cooldown period



RDS, Aurora & ElastiCache

Amazon RDS Overview



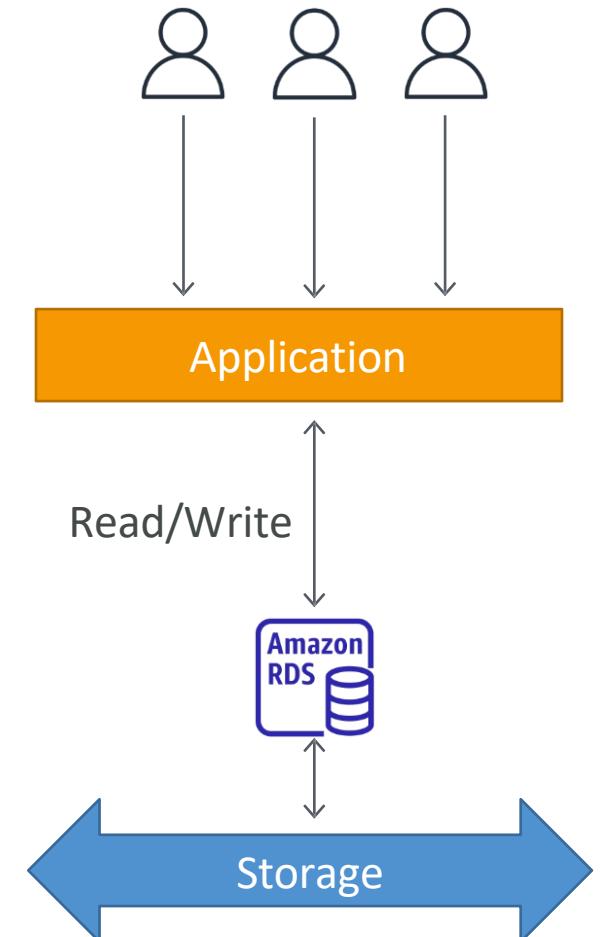
- RDS stands for Relational Database Service
- It's a managed DB service for DB use SQL as a query language.
- It allows you to create databases in the cloud that are managed by AWS
 - Postgres
 - MySQL
 - MariaDB
 - Oracle
 - Microsoft SQL Server
 - IBM DB2
 - Aurora (AWS Proprietary database)

Advantage over using RDS versus deploying DB on EC2

- RDS is a managed service:
 - Automated provisioning, OS patching
 - Continuous backups and restore to specific timestamp (Point in Time Restore)!
 - Monitoring dashboards
 - Read replicas for improved read performance
 - Multi AZ setup for DR (Disaster Recovery)
 - Maintenance windows for upgrades
 - Scaling capability (vertical and horizontal)
 - Storage backed by EBS
- BUT you can't SSH into your instances

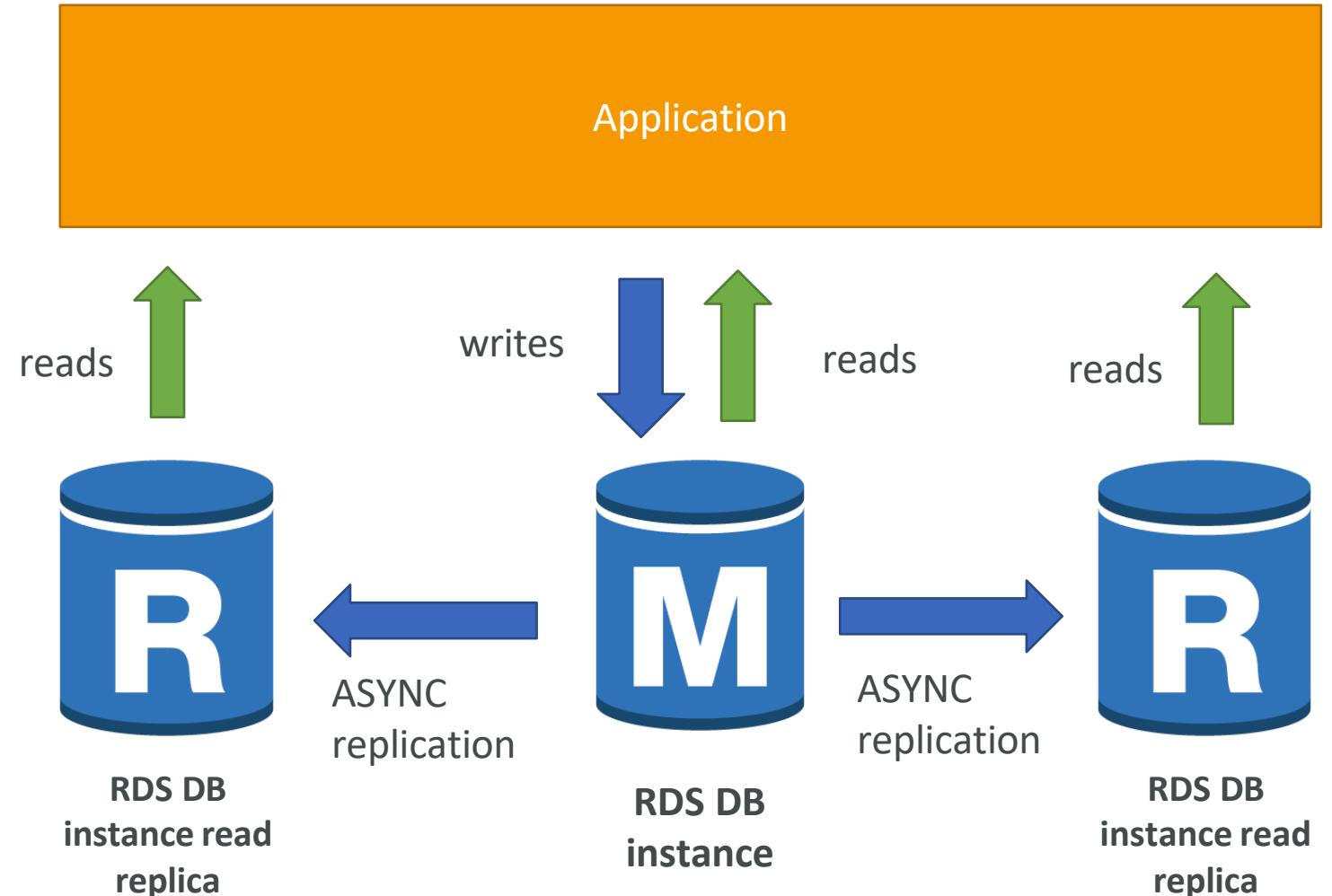
RDS - Storage Auto Scaling

- Helps you increase storage on your RDS DB instance dynamically
- When RDS detects you are running out of free database storage, it scales automatically
- Avoid manually scaling your database storage
- You have to set Maximum Storage Threshold (maximum limit for DB storage)
- Automatically modify storage if:
 - Free storage is less than 10% of allocated storage
 - Low-storage lasts at least 5 minutes
 - 6 hours have passed since last modification
- Useful for applications with unpredictable workloads
- Supports all RDS database engines



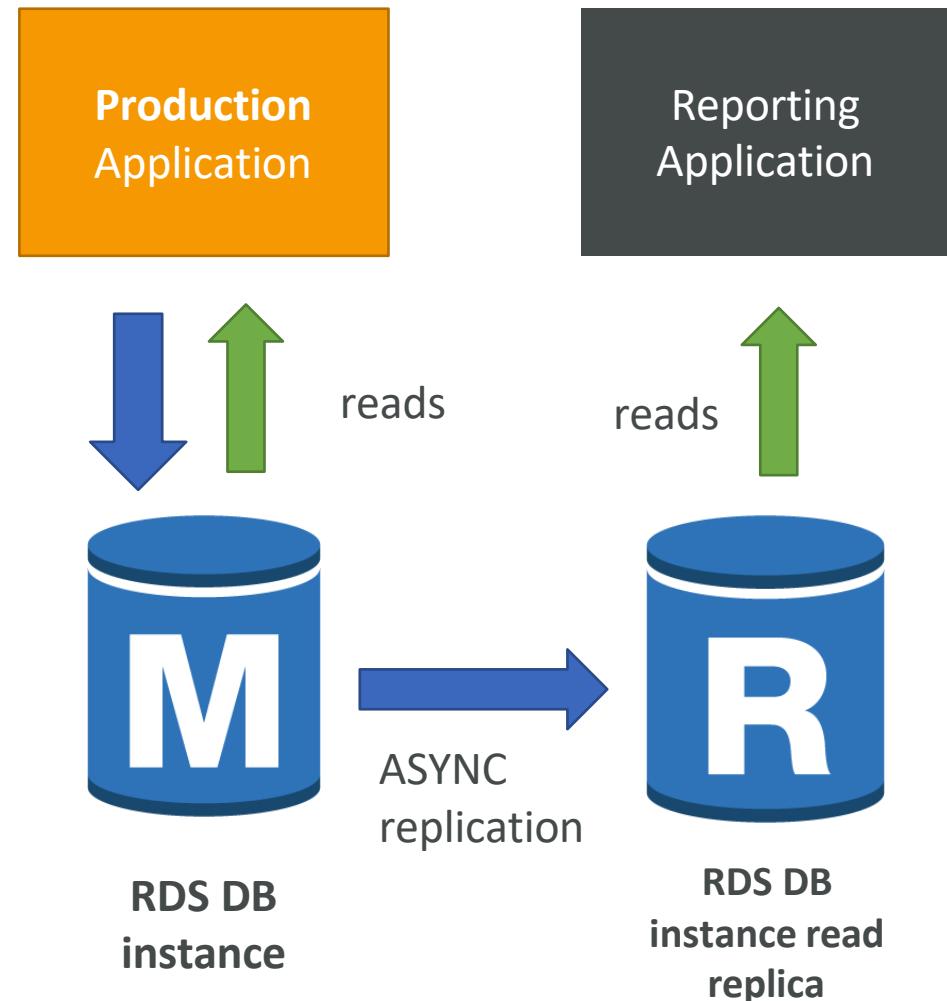
RDS Read Replicas for read scalability

- Up to 15 Read Replicas
- Within AZ, Cross AZ or Cross Region
- Replication is ASYNC, so reads are eventually consistent
- Replicas can be promoted to their own DB
- Applications must update the connection string to leverage read replicas



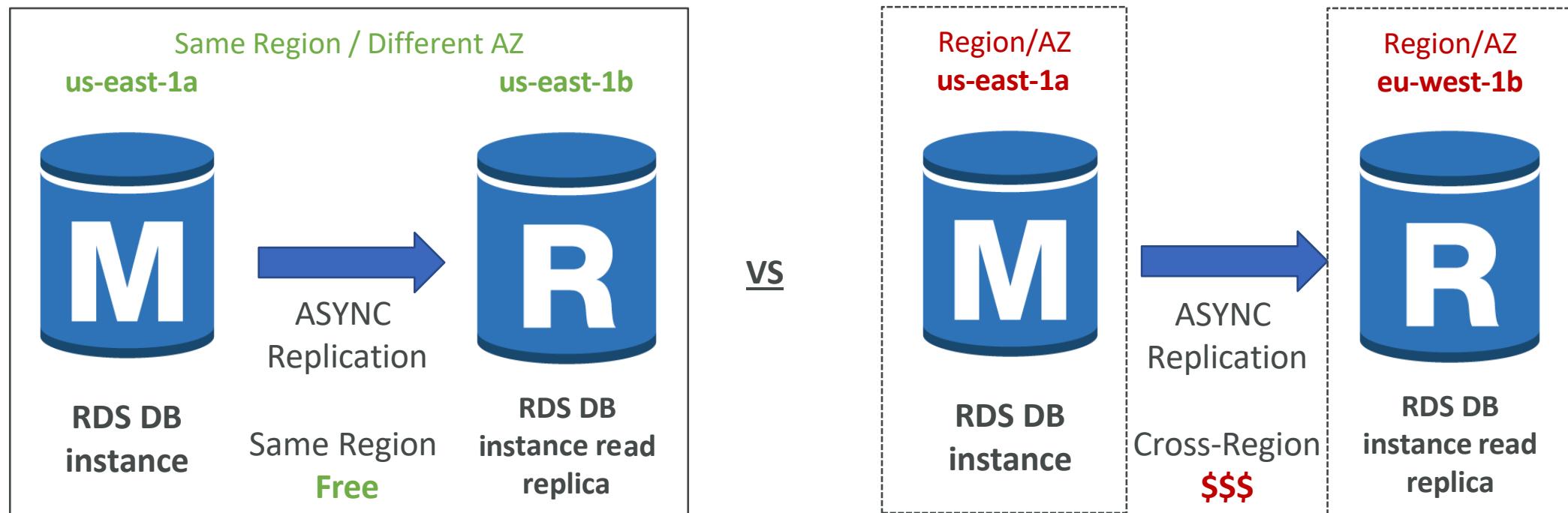
RDS Read Replicas - Use Cases

- You have a production database that is taking on normal load
- You want to run a reporting application to run some analytics
- You create a Read Replica to run the new workload there
- The production application is unaffected
- Read replicas are used for SELECT (=read) only kind of statements (not INSERT, UPDATE, DELETE)



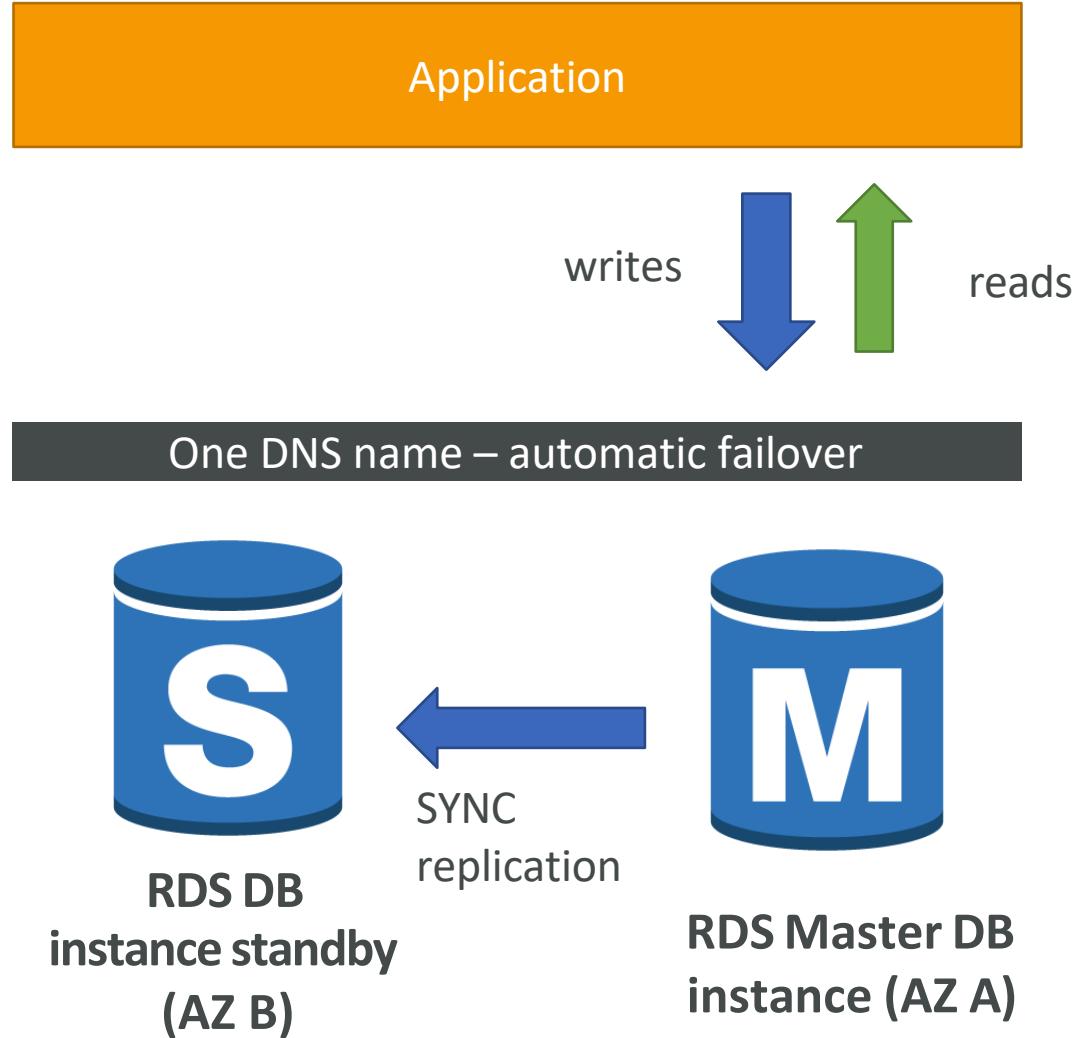
RDS Read Replicas - Network Cost

- In AWS there's a network cost when data goes from one AZ to another
- For RDS Read Replicas within the same region, you don't pay that fee



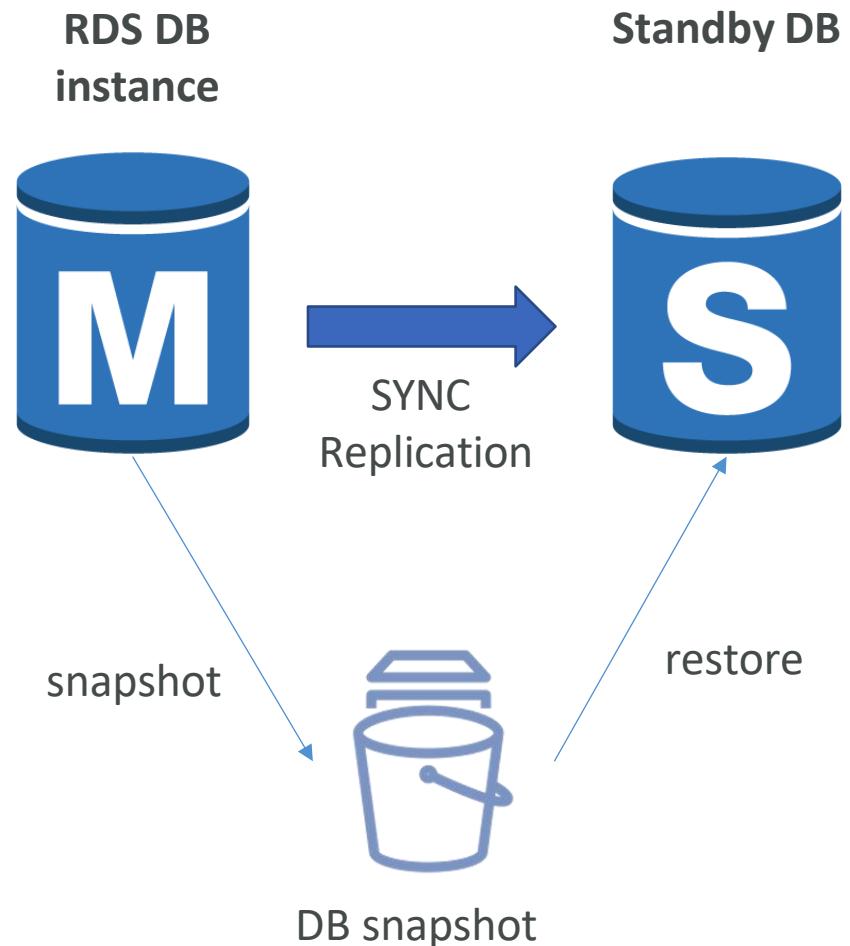
RDS Multi AZ (Disaster Recovery)

- SYNC replication
- One DNS name - automatic app failover to standby
- Increase availability
- Failover in case of loss of AZ, loss of network, instance or storage failure
- No manual intervention in apps
- Not used for scaling
- Note:The Read Replicas be setup as Multi AZ for Disaster Recovery (DR)



RDS - From Single-AZ to Multi-AZ

- Zero downtime operation (no need to stop the DB)
- Just click on “modify” for the database
- The following happens internally:
 - A snapshot is taken
 - A new DB is restored from the snapshot in a new AZ
 - Synchronization is established between the two databases



RDS Custom

- Managed Oracle and Microsoft SQL Server Database with OS and database customization
- RDS: Automates setup, operation, and scaling of database in AWS
- Custom: access to the underlying database and OS so you can
 - Configure settings
 - Install patches
 - Enable native features
 - Access the underlying EC2 Instance using SSH or SSM Session Manager
- De-activate Automation Mode to perform your customization, better to take a DB snapshot before
- RDS vs. RDS Custom
 - RDS: entire database and the OS to be managed by AWS
 - RDS Custom: full admin access to the underlying OS and the database

