AWS Cloud Practitioner Certification Bootcamp

Week - 2

Session 2 - Amazon S3 & Ec2

22nd January, Saturday
7:30 PM to 9:00 PM IST

WOMEN IN
BIG DATA
BANGLADESH

aws USER GROUP
BANGLADESH

AWS
User Groups

Speakers

Sanchit Jain

Lead Architect - AWS at Quantiphi

AWS APN Ambassador

Agenda

- Re-cap of Last session
- Amazon S3
- Amazon Compute (EC2)

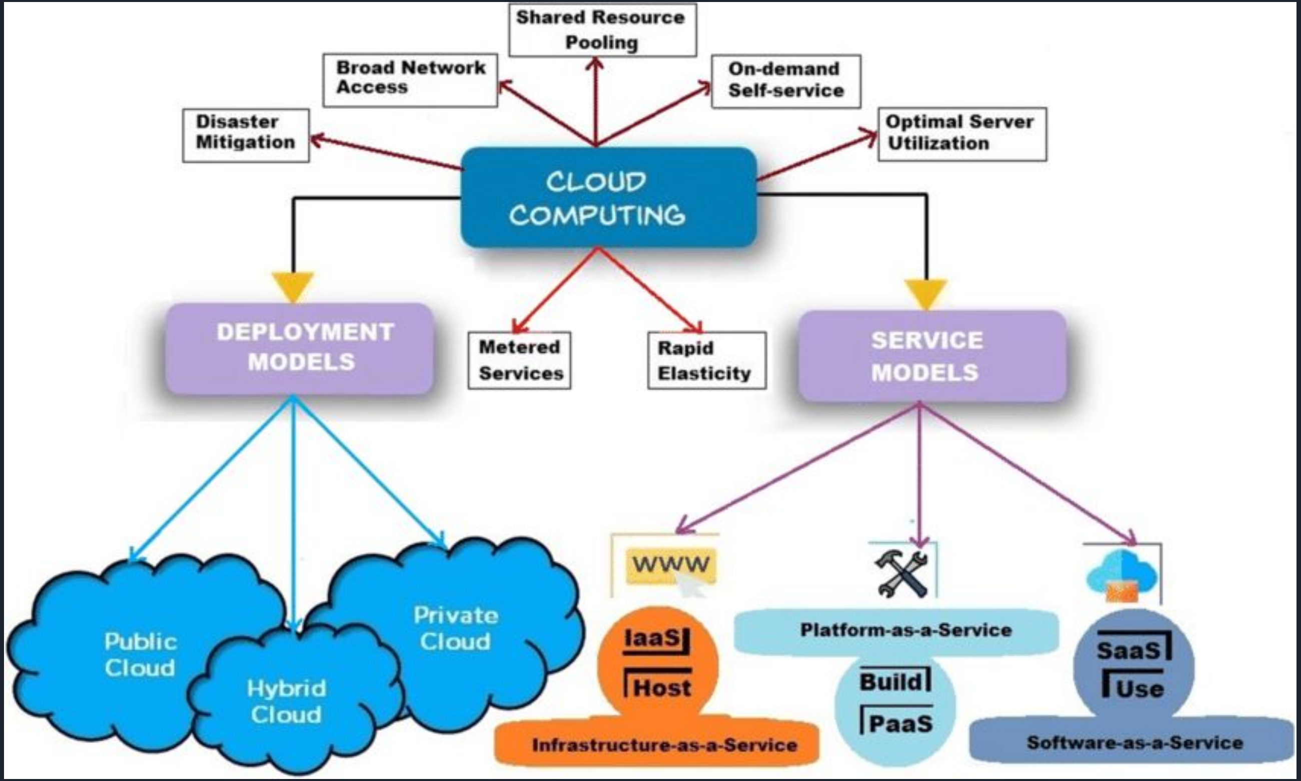
AWS
User Groups

Re-cap of Last session

Re-cap

AWS
User Groups

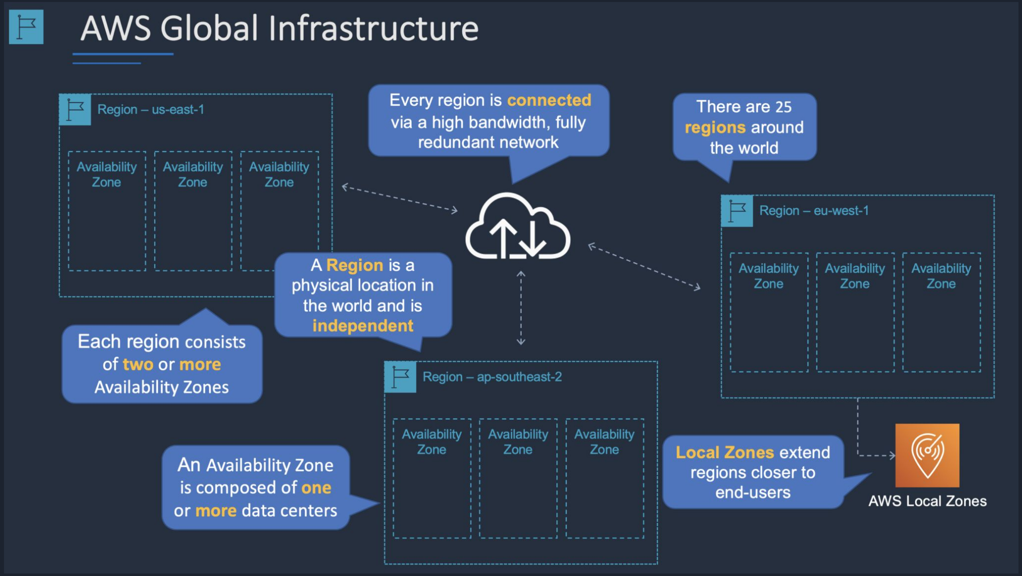
Cloud Computing



Introduction to AWS

- Amazon Web Services (AWS) is a secure cloud services platform offering compute power, database storage, analytics, application and deployment services that help organizations move faster, lower IT costs, and scale applications.

- AWS provides services from dozens of data centers spread across availability zones (AZs) in regions across the world.



AWS Global Infrastructure

Every region is **connected** via a high bandwidth, fully redundant network

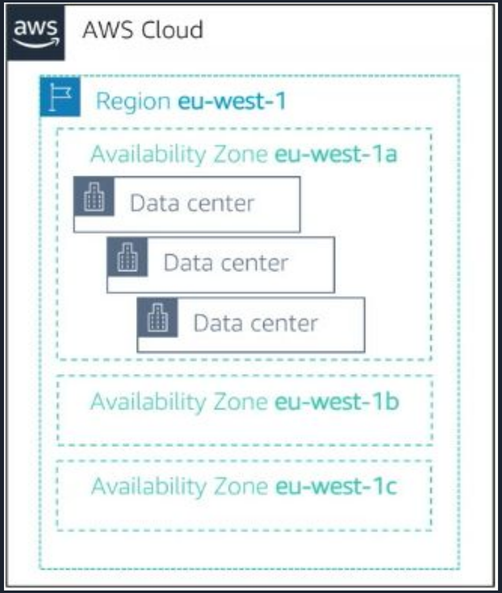
There are 25 **regions** around the world

A **Region** is a physical location in the world and is **independent**

Each region consists of **two** or **more** Availability Zones

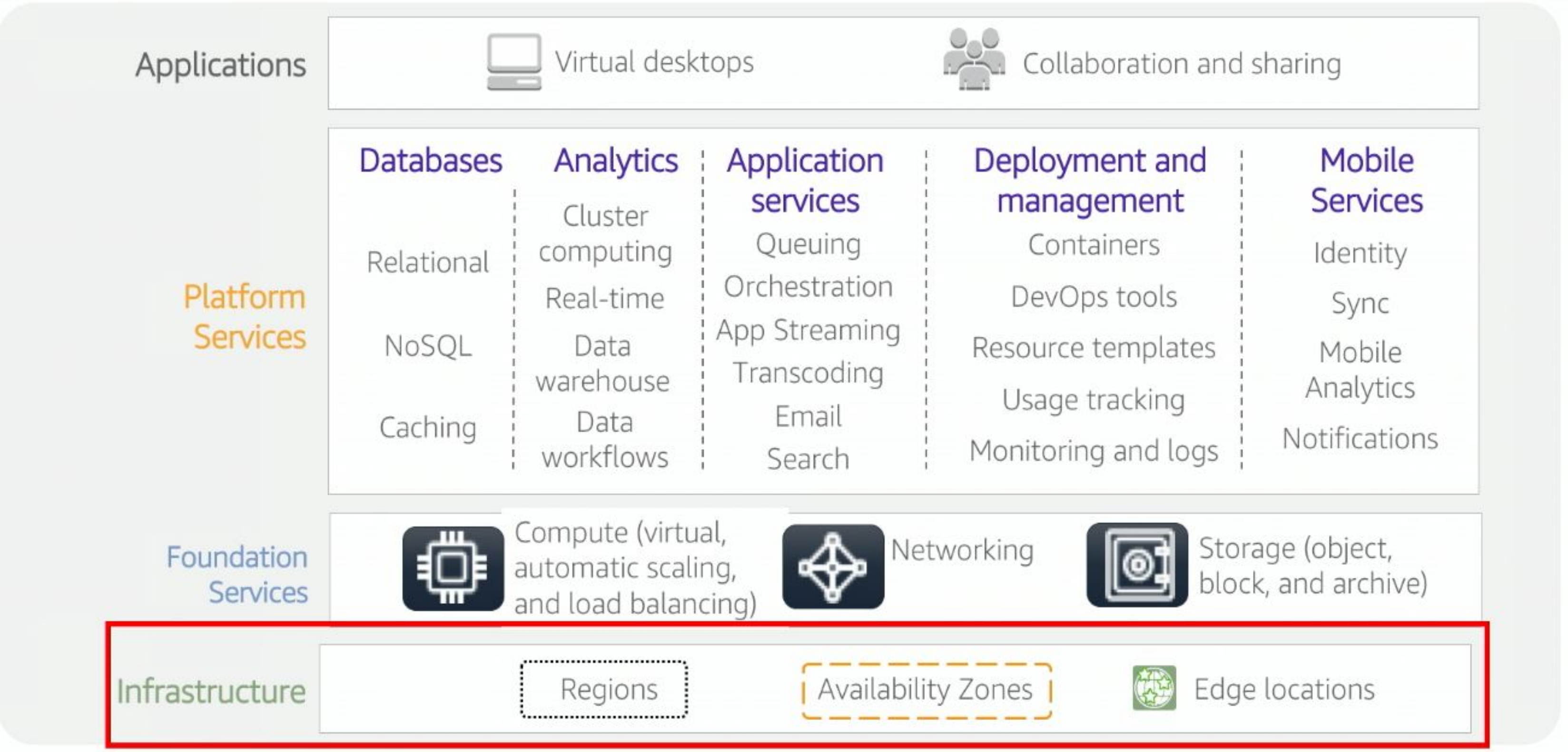
An Availability Zone is composed of **one** or **more** data centers

Local Zones extend regions closer to end-users



aws
AWS Cloud
Region eu-west-1
Availability Zone eu-west-1a
Data center
Data center
Data center
Availability Zone eu-west-1b
Availability Zone eu-west-1c

AWS Services



Applications

Virtual desktops

Collaboration and sharing

Platform Services

Databases

Relational

NoSQL

Caching

Analytics

Cluster computing

Real-time

Data warehouse

Data workflows

Application services

Queuing

Orchestration

App Streaming

Transcoding

Email

Search

Deployment and management

Containers

DevOps tools

Resource templates

Usage tracking

Monitoring and logs

Mobile Services

Identity

Sync

Mobile Analytics

Notifications

Foundation Services

Compute (virtual, automatic scaling, and load balancing)

Networking

Storage (object, block, and archive)

Infrastructure

Regions

Availability Zones

Edge locations

Amazon S3

Amazon S3 Overview

- Infinitely scaling storage
- Unlimited storage space & pay-as-you-use model.
- Amazon S3 allows people to store objects (files) in “buckets” (directories). S3 resources for e.g. buckets and objects are private by default
- Event notifications for specific actions, can send alerts or trigger actions, and it can be sent to:
  - SNS Topics.
  - SQS Queue.
  - Lambda functions.
  - Need to configure SNS/SQS/Lambda before S3.
  - No extra charges from S3 but you pay for SNS, SQS and Lambda.

Amazon S3 Overview - Buckets

- Buckets are defined at the region level but must have a globally unique name
- Bucket Naming convention - No uppercase, underscore
- Naming convention
  - No uppercase
  - No underscore
  - 3-63 characters long
  - Not an IP
  - Must start with lowercase letter or number

Amazon S3 Overview - Objects

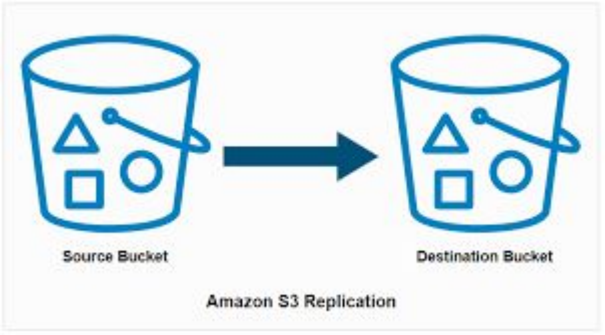
- The key is composed of prefix + object name s3://my-bucket/my_folder1/another_folder/my_file.txt
- Objects consist of object data, metadata, and others
  - Key
  - Value
  - Metadata
  - Version ID (if versioning is enabled)
  - Access Control Information
- S3 objects allow two kinds of metadata - System metadata and User-defined metadata(“x-amz-meta”)
- Metadata cannot be modified after the object is uploaded
- Each object can be up to 5 TB in size

Amazon S3 - Versioning

- Version Object
- Enabled at the bucket level
- Protect against unintended deletes (ability to restore a version)
- Old versions count as billable size until they are permanently deleted
- Cross Region Replication requires versioning to be enabled on the source and destination buckets.
- Notes:
  - Any file that is not versioned prior to enabling versioning will have version “null”
  - Suspending versioning does not delete the previous versions

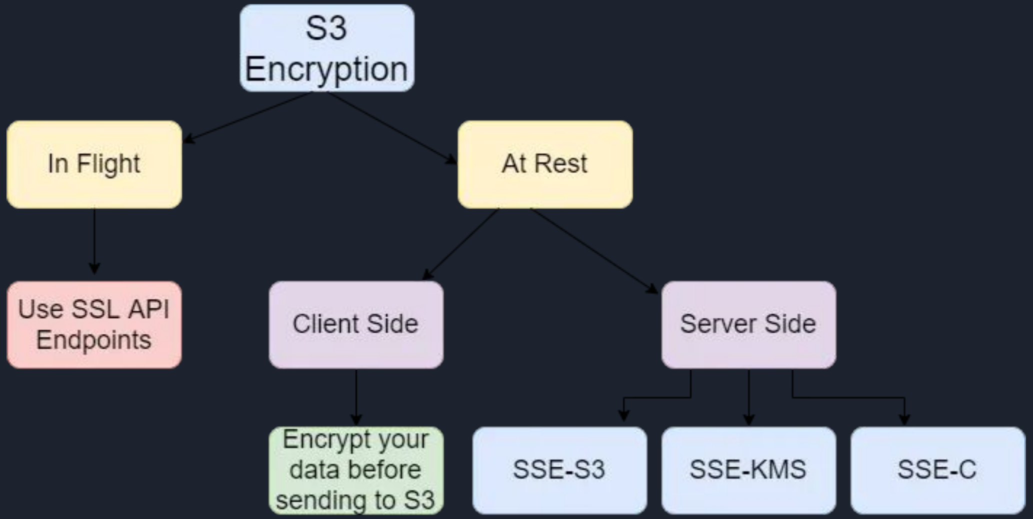
S3 Replication

- Must enable versioning in source and destination
- Cross Region Replication (CRR)
- Same Region Replication (SRR)
- Buckets can be in different accounts
- Copying is asynchronous
- CRR - Use cases: compliance, lower latency access, replication across accounts
- SRR – Use cases: log aggregation, live replication between production and test accounts

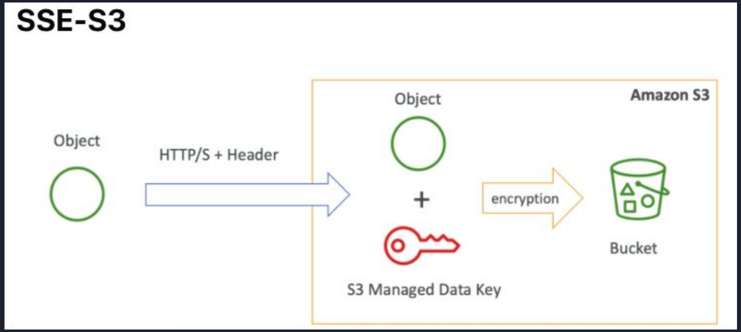


S3 Encryption for Objects

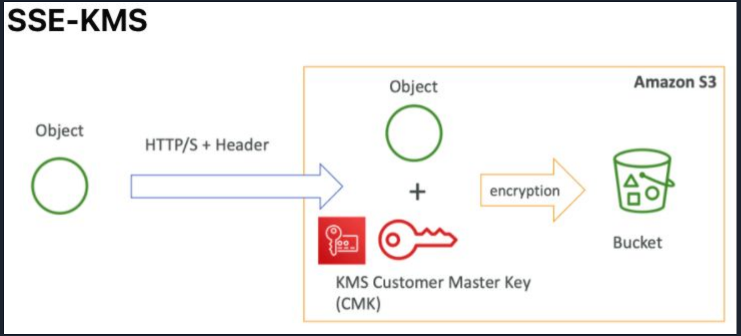
- AWS S3 Encryption supports both data at rest and data in transit encryption.
- Data in-transit
- Data at Rest
  - Server-Side Encryption
  - Client-Side Encryption
- The encryption process, the encryption keys, and related tools are managed by the user.



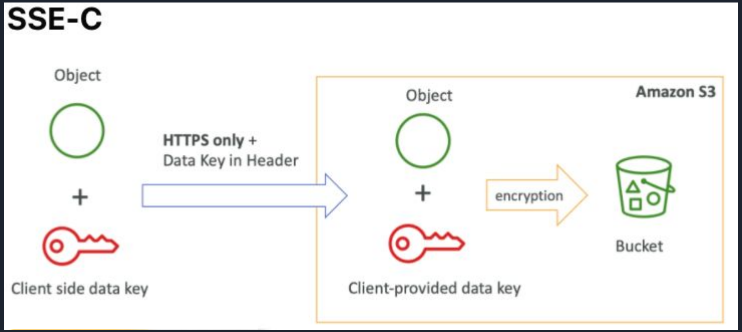
S3 Encryption for Objects



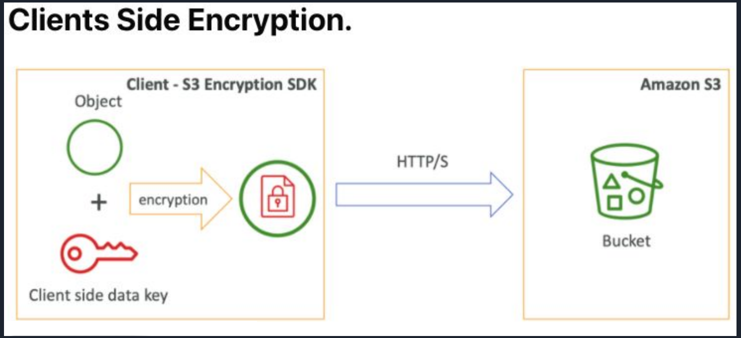
SSE-S3



SSE-KMS



SSE-C

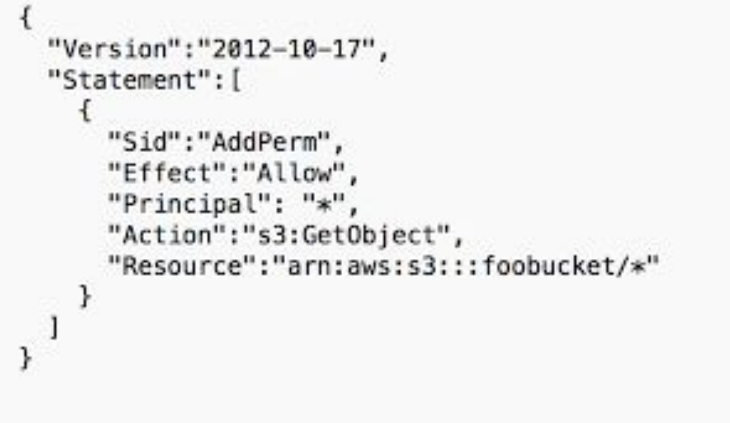


Clients Side Encryption.

S3 Bucket Policies

JSON based policies

- Resources: buckets and objects
- Actions: Set of API to Allow or Deny
- Effect: Allow / Deny
- Principal:The account or user to apply the policy to



```
{
  "Version":"2012-10-17",
  "Statement":[
    {
      "Sid":"AddPerm",
      "Effect":"Allow",
      "Principal": "*",
      "Action":"s3:GetObject",
      "Resource":"arn:aws:s3:::foobucket/*"
    }
  ]
}
```

Use S3 bucket for policy to:

- Grant public access to the bucket
- Force objects to be encrypted at upload
- Grant access to another account (Cross Account)

S3 Websites

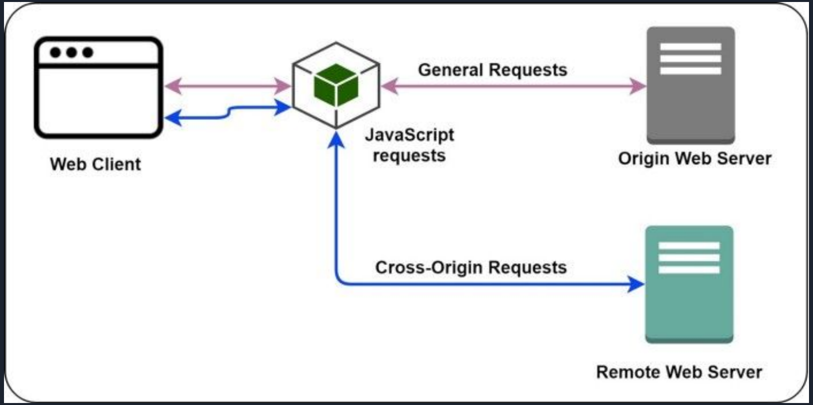
- S3 can be used to host static websites.
- Cannot use dynamic content such as PHP, .Net etc.
- Automatically scales.
- You can use a custom domain name with S3 using a Route 53 Alias record.
- The website URL will be:
  - <bucket-name>.s3-website-<AWS-region>.amazonaws.com

    OR

  - <bucket-name>.s3-website.<AWS-region>.amazonaws.com
- If you get a 403 (Forbidden) error, make sure the bucket policy allows public reads!

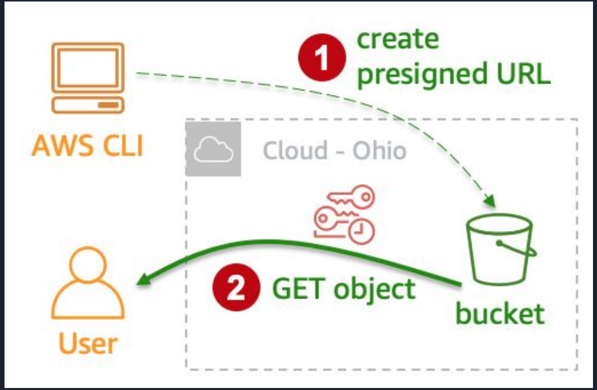
S3 CORS

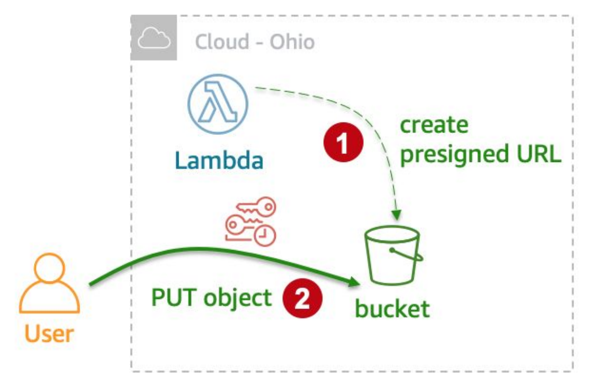
- If a client does a cross-origin request on our S3 bucket, we need to enable the correct CORS headers
- You can allow for a specific origin or for * (all origins)
- Cross-origin HTTP requests can be made to:
  - A different domain (for example, from example.com to amazondomains.com)
  - A different subdomain (for example, from example.com to petstore.example.com)
  - A different port (for example, from example.com to example.com:10777)
  - A different protocol (for example, from https://example.com to http://example.com)



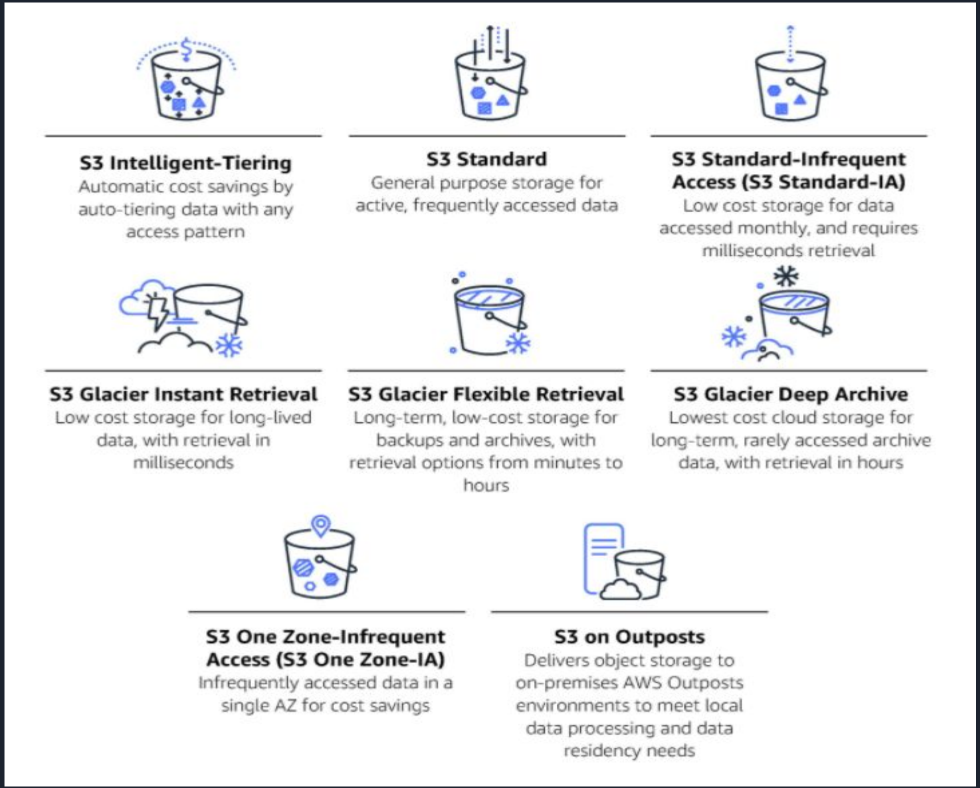
S3 Pre-Signed URLs

- Can generate pre-signed URLs using SDK or CLI
- For downloads (easy, can use the CLI)
- For uploads (harder, must use the SDK)
- Valid for a default of 3600 seconds, can change timeout with --expires-in [TIME_BY_SECONDS] argument
- Users given a pre-signed URL inherit the permissions of the person who generated the URL for GET / PUT





S3 Storage Classes



S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

AWS
User Groups

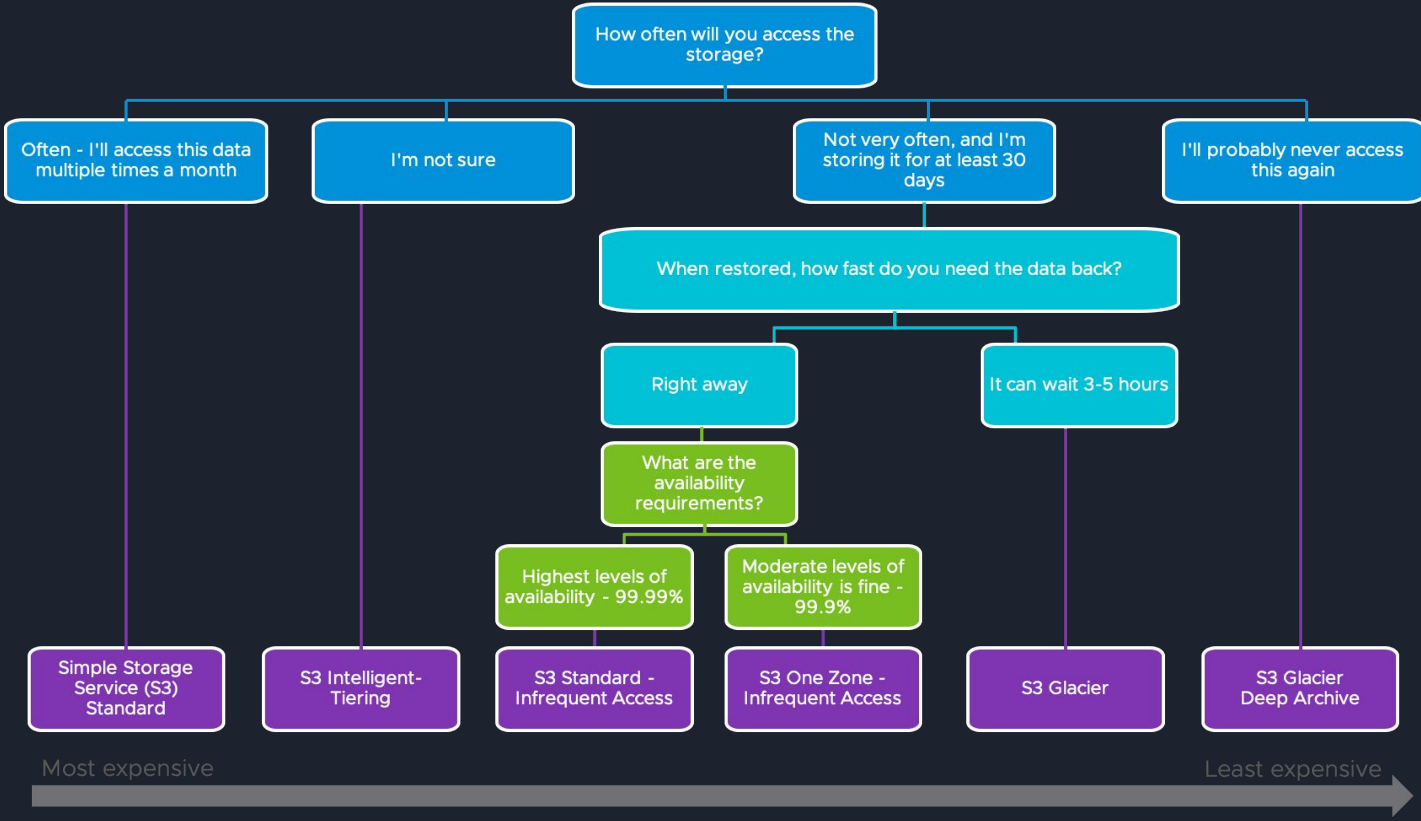
https://aws.amazon.com/s3/storage-classes-infographic/

S3 Storage Classes Comparison

	S3 Standard	S3 Intelligent - Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for availability	99.99%	99.90%	99.90%	99.50%	99.90%	99.99%	99.99%
Availability SLA	99.90%	99%	99%	99%	99%	99.00%	99.90%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128 KB	128 KB	128 KB	40 KB	40 KB
Minimum storage duration charge	N/A	N/A	30 days	30 days	90 days	90 days	180 days
Retrieval charge	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	milliseconds	minutes or hours	hours
Storage type	Object	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes	Yes

https://aws.amazon.com/s3/storage-classes/

How to choose S3 Storage Class?

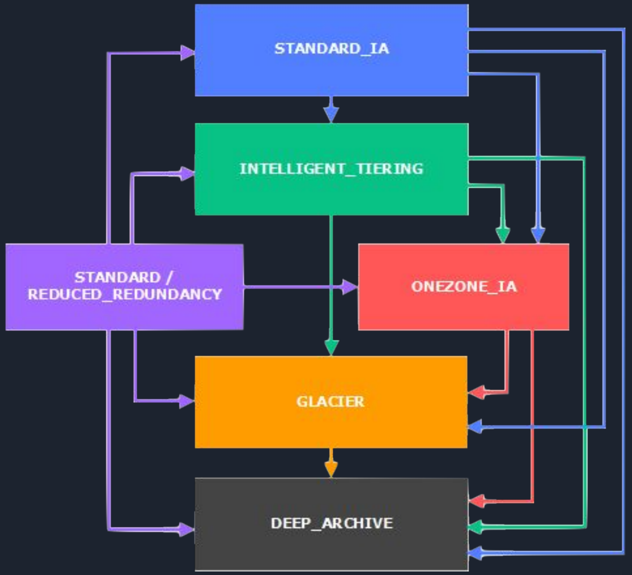

How often will you access the storage?
Often - I'll access this data multiple times a month
I'm not sure
Not very often, and I'm storing it for at least 30 days
I'll probably never access this again
When restored, how fast do you need the data back?
Right away
It can wait 3-5 hours
What are the availability requirements?
Highest levels of availability - 99.99%
Moderate levels of availability is fine - 99.9%
Simple Storage Service (S3) Standard
S3 Intelligent-Tiering
S3 Standard - Infrequent Access
S3 One Zone - Infrequent Access
S3 Glacier
S3 Glacier Deep Archive
Most expensive
Least expensive

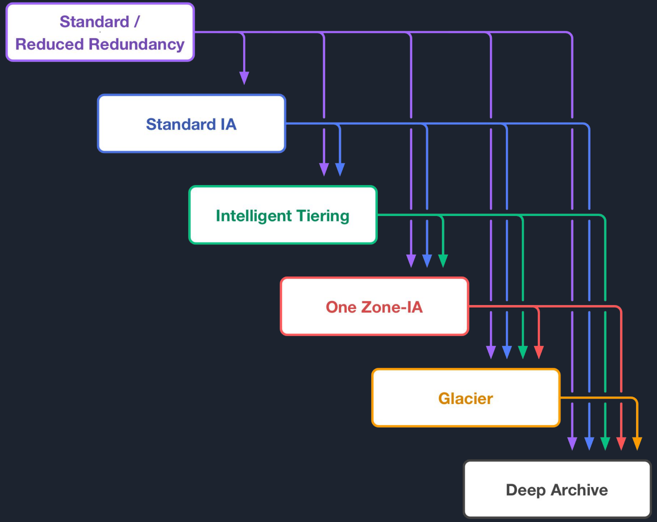
Source: CloudHealth by VMware

S3 – Moving between storage classes

You can transition objects between storage classes via S3 Lifecycle Rules

- Moving objects can be automated using a lifecycle configuration
- We can define Transition actions or Expiration actions in S3 Lifecycle Rules



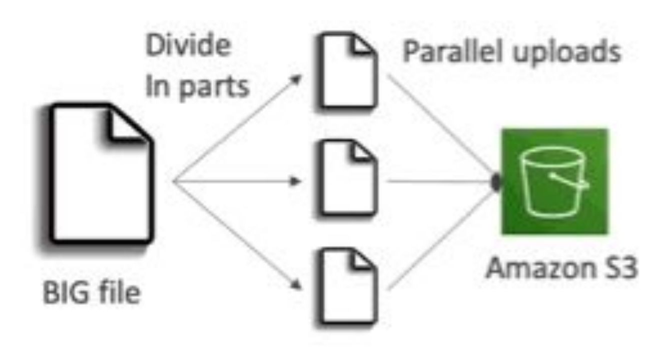


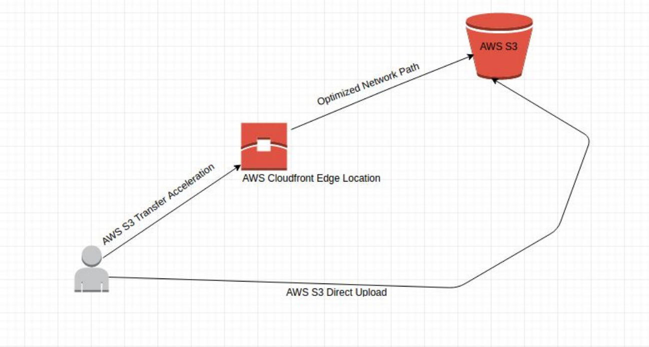
AWS User Groups

S3 Performance

- Multi-Part upload
  - recommended for files > 100MB, must use for files > 5GB
  - Can help parallelize uploads (speed up transfers)

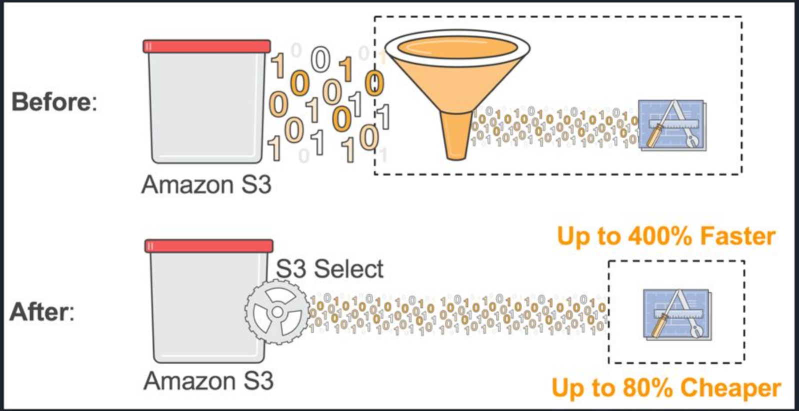
- S3 Transfer Acceleration
  - Increase transfer speed by transferring file to an AWS edge location which will forward the data to the S3 bucket in the target region
  - Compatible with multi-part upload

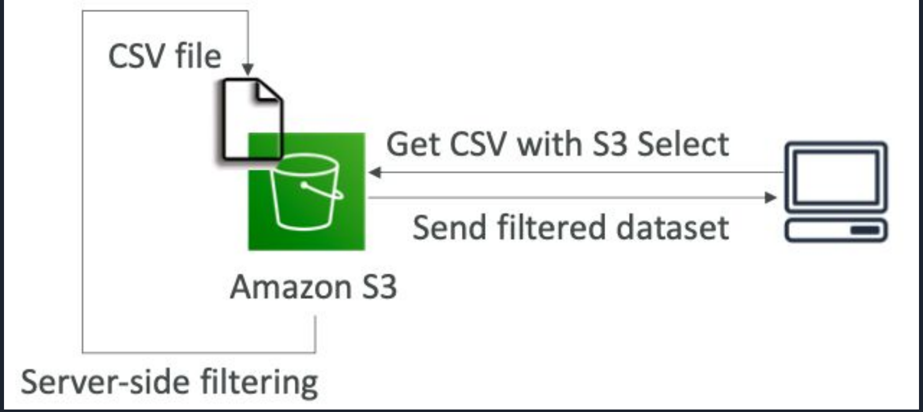




S3 Select & Glacier Select

- Retrieve less data using SQL by performing server side filtering
- Can filter by rows & columns (simple SQL statements)
- Less network transfer, less CPU cost client-side

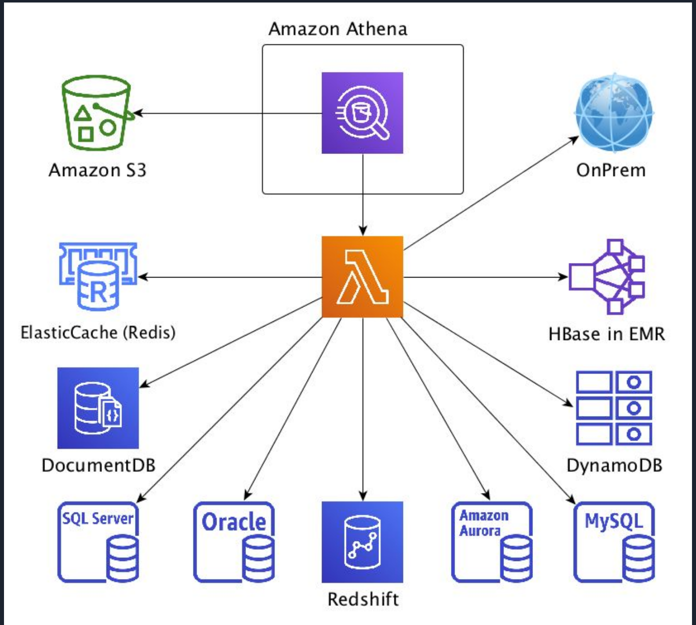




https://aws.amazon.com/blogs/aws/s3-glacier-select/

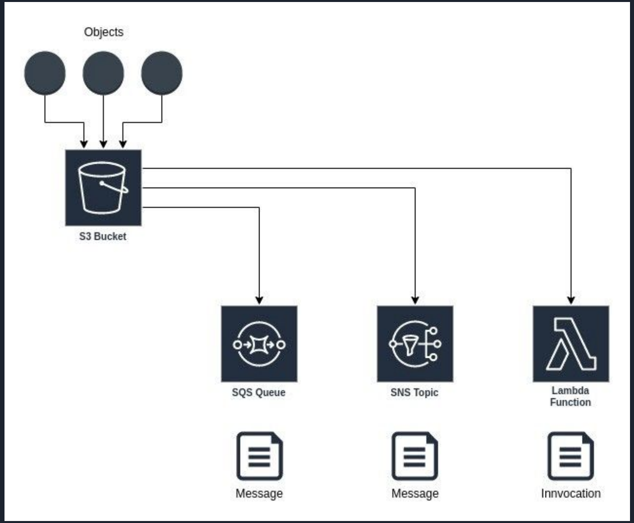
Amazon Athena

- Serverless query service to perform analytics against S3 objects
- Uses standard SQL language to query the files
- Supports CSV,JSON,ORC,Avro,and Parquet(builtonPresto)
- Pricing: $5.00 per TB of data scanned
- Use compressed or columnar data for cost-savings (less scan)
- Use cases: Business intelligence / analytics, analyze & query VPC Flow Logs, ELB Logs, CloudTrail trails, etc...
- Exam Tip: analyze data in S3 using serverless SQL, use Athena



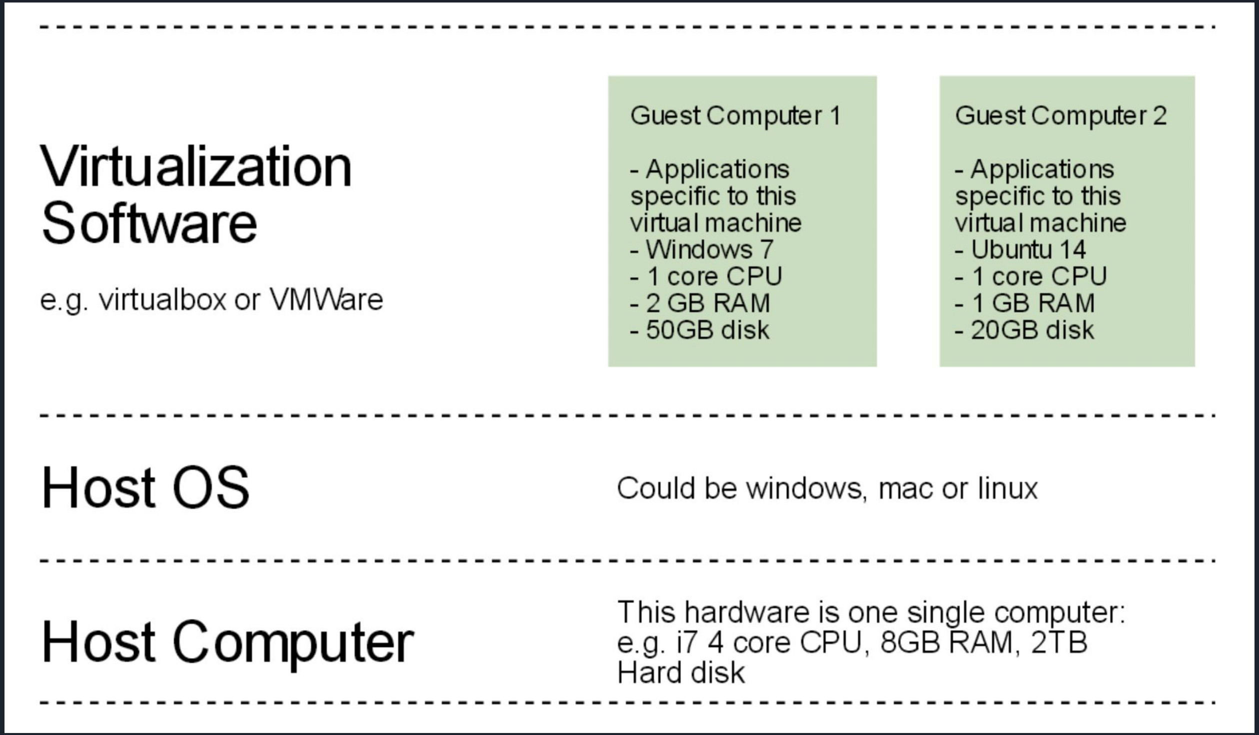
S3 Event Notifications

- Amazon S3 event notifications can be sent in response to actions in Amazon S3 like PUTs, POSTs, COPYs, or DELETEs.
- You can configure notifications to be filtered by the prefix and suffix of the key name of objects.
- Amazon S3 can publish notifications for the following events:
  - New object created events.
  - Object removal events.
  - Restore object events.
  - Reduced Redundancy Storage (RRS) object lost events.
  - Replication events.



Amazon Compute(EC2)

Virtualisation On-prem



Virtualization Software

e.g. virtualbox or VMWare

Guest Computer 1

- Applications specific to this virtual machine
- Windows 7
- 1 core CPU
- 2 GB RAM
- 50GB disk

Guest Computer 2

- Applications specific to this virtual machine
- Ubuntu 14
- 1 core CPU
- 1 GB RAM
- 20GB disk

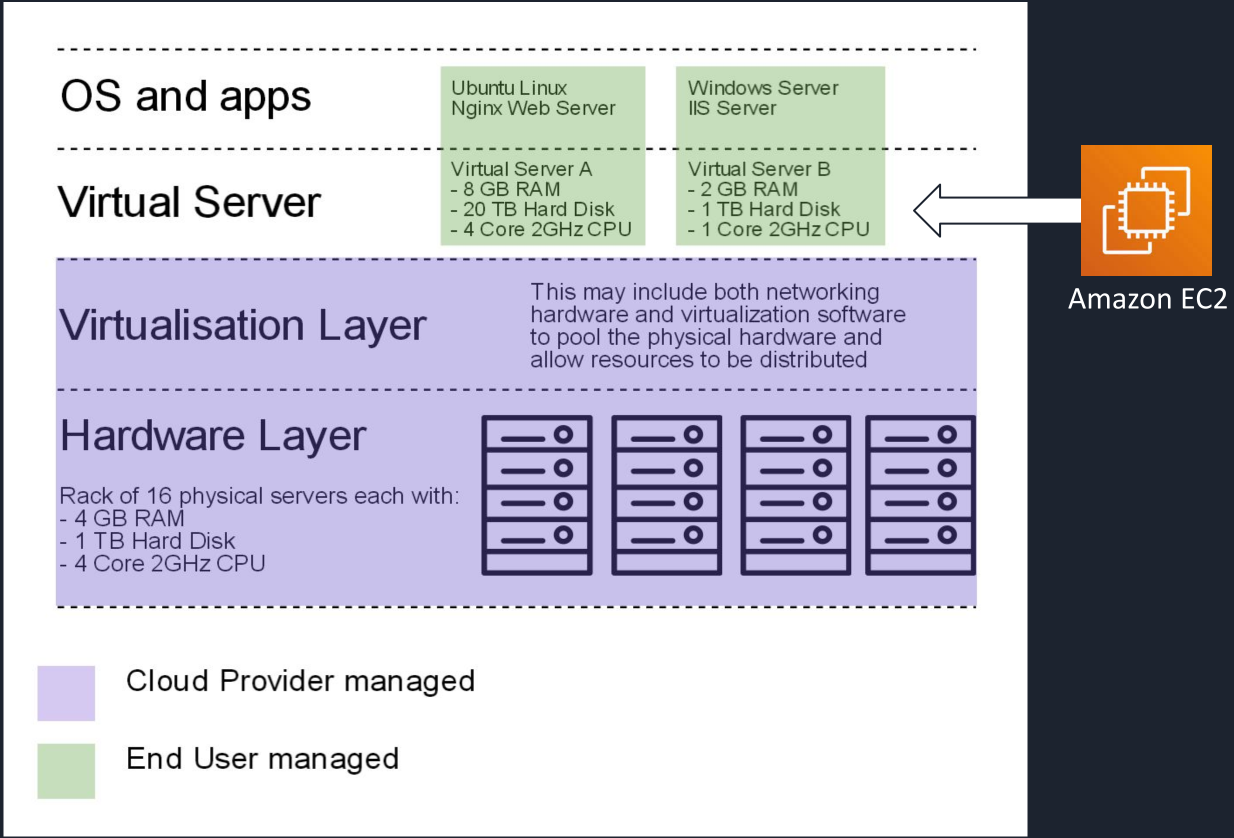
Host OS

Could be windows, mac or linux

Host Computer

This hardware is one single computer: e.g. i7 4 core CPU, 8GB RAM, 2TB Hard disk

Amazon EC2



OS and apps

Ubuntu Linux
Nginx Web Server

Windows Server
IIS Server

Virtual Server

Virtual Server A
- 8 GB RAM
- 20 TB Hard Disk
- 4 Core 2GHz CPU

Virtual Server B
- 2 GB RAM
- 1 TB Hard Disk
- 1 Core 2GHz CPU

Amazon EC2

Virtualisation Layer

This may include both networking hardware and virtualization software to pool the physical hardware and allow resources to be distributed

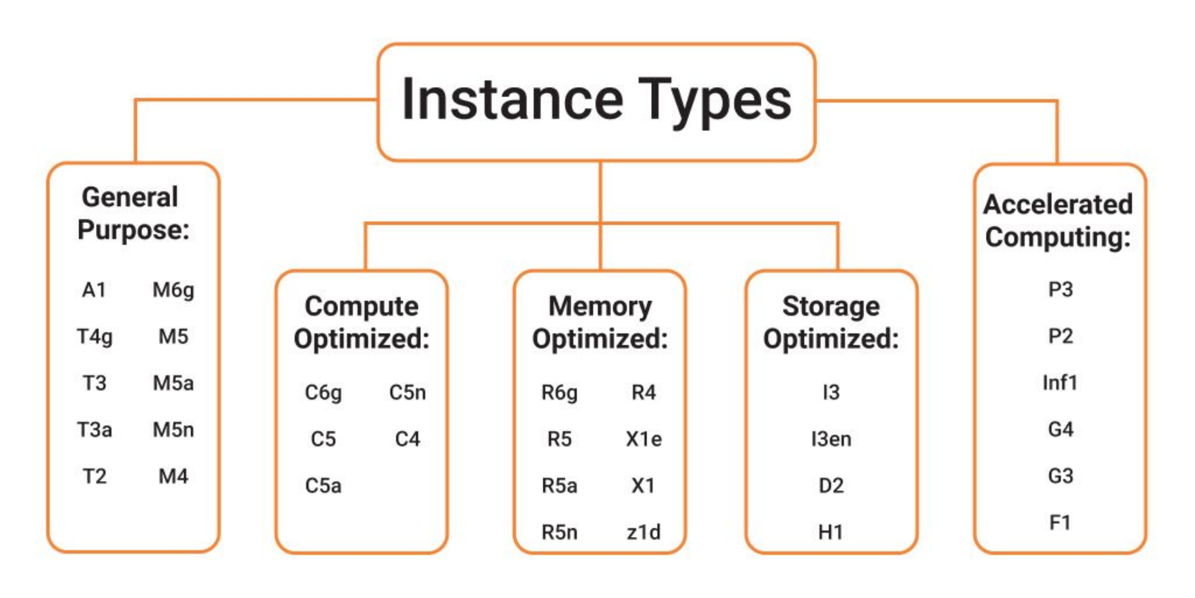
Hardware Layer

Rack of 16 physical servers each with:
- 4 GB RAM
- 1 TB Hard Disk
- 4 Core 2GHz CPU

🟪 Cloud Provider managed

🟩 End User managed

Amazon EC2 Instance Families



Instance Types

General Purpose:

A1
T4g
T3
T3a
T2
M6g
M5
M5a
M5n
M4

Compute Optimized:

C6g
C5
C5a
C5n
C4

Memory Optimized:

R6g
R5
R5a
R5n
R4
X1e
X1
z1d

Storage Optimized:

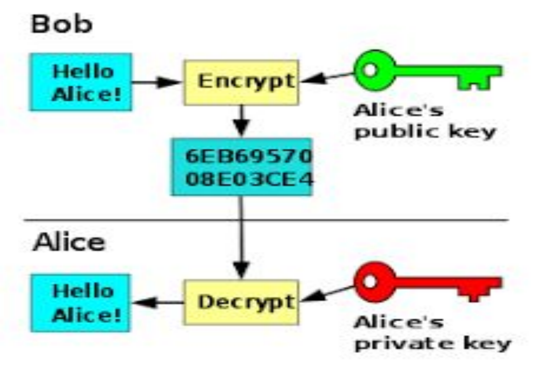
I3
I3en
D2
H1

Accelerated Computing:

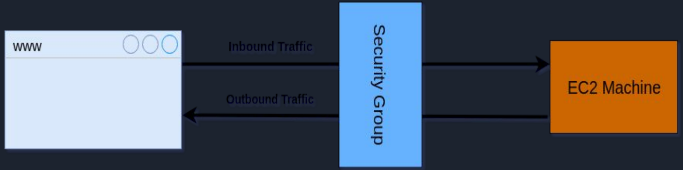
P3
P2
Inf1
G4
G3
F1

Amazon EC2 Security

- EC2 Key Pairs
  - When you launch an instance, you specify the key pair which you require to use.
  - At the boot time, the public key content is placed on the instance in an entry within ~/.ssh/authorized_keys
  - To log in to your instance, you must specify the private key when you connect to the instance

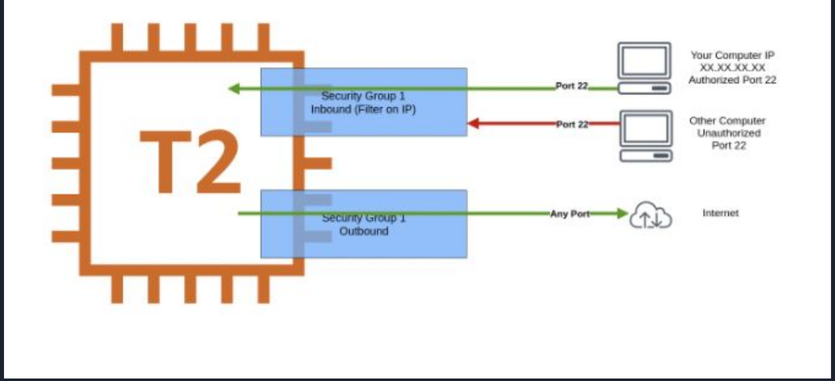


- Security groups
  - Security groups are virtual firewall that controls the traffic for an instance/RDS.
  - When you launch an instance, you can specify one or more security groups
  - You can add rules to each security group that allow traffic to or from its associated instances (Inbound & Outbound Rules)



Amazon EC2 Security

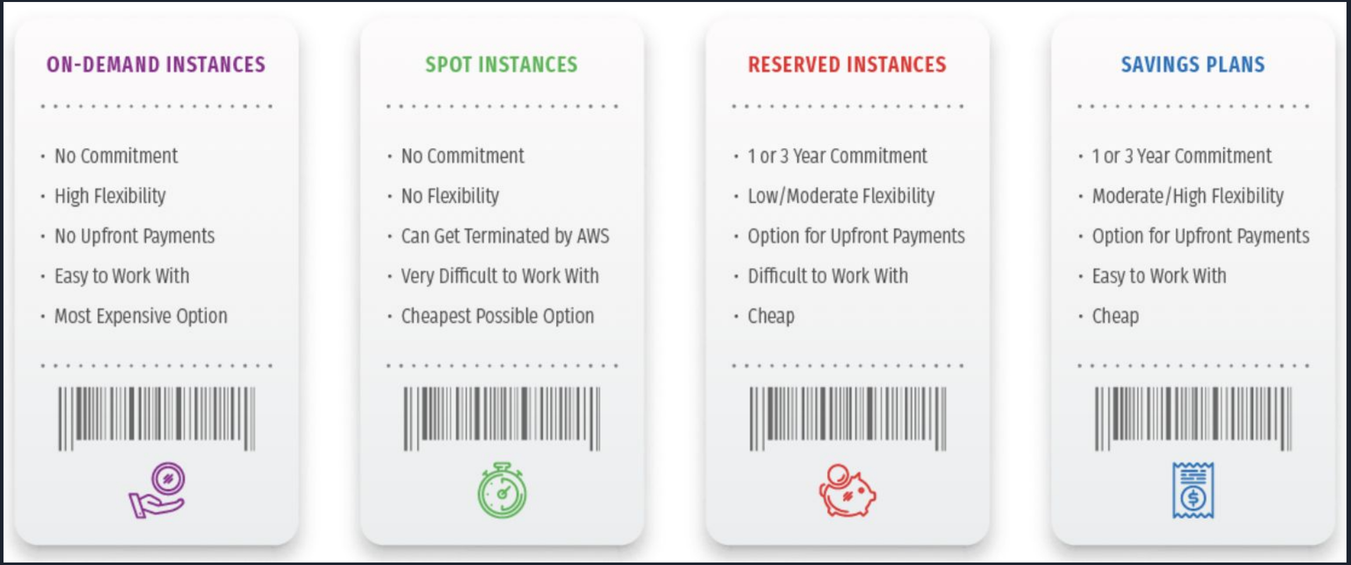
- Security groups - Classic Ports to know
  - 22 = SSH (Secure Shell) - log into a Linux instance
  - 21 = FTP (File Transfer Protocol) – upload files into a file share
  - 22 = SFTP (Secure File Transfer Protocol) – upload files using SSH
  - 80 = HTTP – access unsecured websites
  - 443 = HTTPS – access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



Amazon EC2 Security

- Rules for AWS Security Groups
  - Security Groups should avoid having large port ranges. This increases the attack surface and increases vulnerability of your EC2 instances.
  - Limit outbound access from ports to specific ports or other destinations.
  - It is good to maintain one security group for SSH Access to your instances since SSH is a critical access.
  - While working with database instances such as RDS, don’t allow unrestricted access to the RDS. Doing so increases the risk of brute-force login attacks.
  - Delete unused security groups as soon as possible. It is a good practice to have regular clean-up exercise on your AWS project environment to ensure there are no unused security groups.
  - Restrict access to security group modification or creation using appropriate IAM policies.

EC2 Instance Billing Model



ON-DEMAND INSTANCES	SPOT INSTANCES	RESERVED INSTANCES	SAVINGS PLANS
No Commitment	No Commitment	1 or 3 Year Commitment	1 or 3 Year Commitment
High Flexibility	No Flexibility	Low/Moderate Flexibility	Moderate/High Flexibility
No Upfront Payments	Can Get Terminated by AWS	Option for Upfront Payments	Option for Upfront Payments
Easy to Work With	Very Difficult to Work With	Difficult to Work With	Easy to Work With
Most Expensive Option	Cheapest Possible Option	Cheap	Cheap

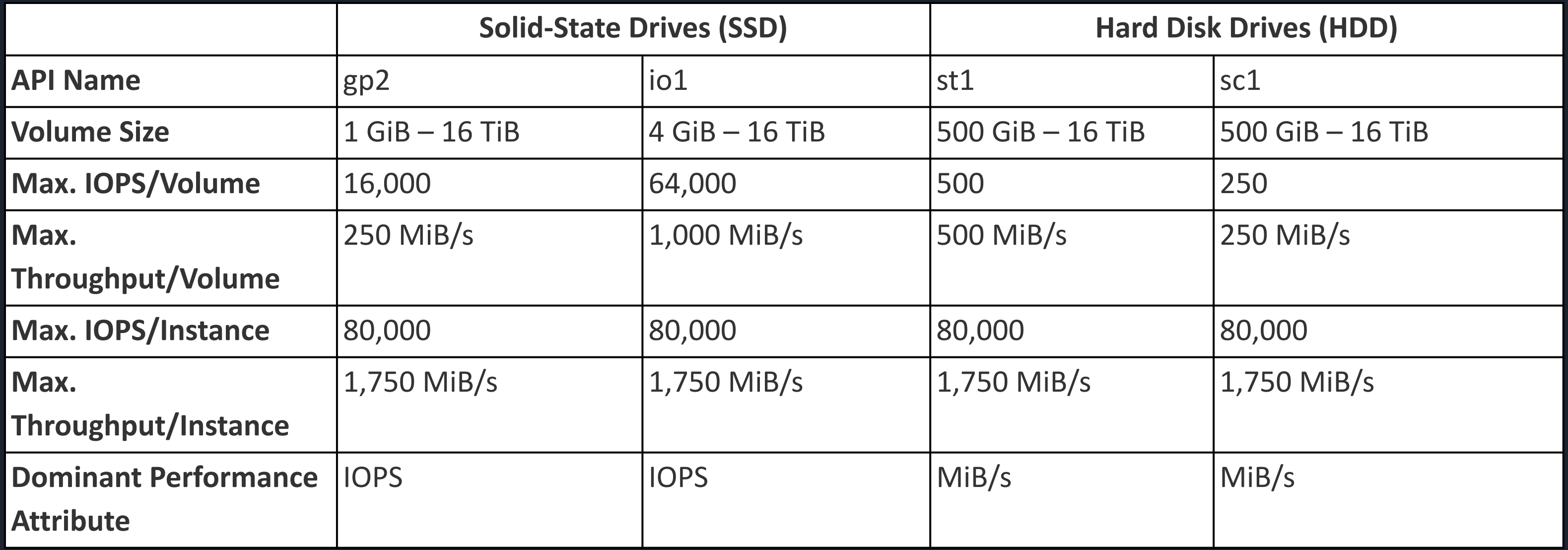
Which purchasing option is right for me?

- On demand: book a OLA cab, whenever you want a ride
- Reserved: outstation ride, reservation for a long time, we may get a good discount.
- Spot instances: shared cab/pool cab, depends on the availability and route. You can get kicked out at any time
- Dedicated Hosts: Your own car

What is EBS?

- 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, and are bound to a specific availability zone
- It's a network drive (i.e. not a physical drive)
- It's locked to an Availability Zone (AZ)
- Have a provisioned capacity (size in GBs, and IOPS)
- Controls the EBS behaviour when an EC2 instance terminates

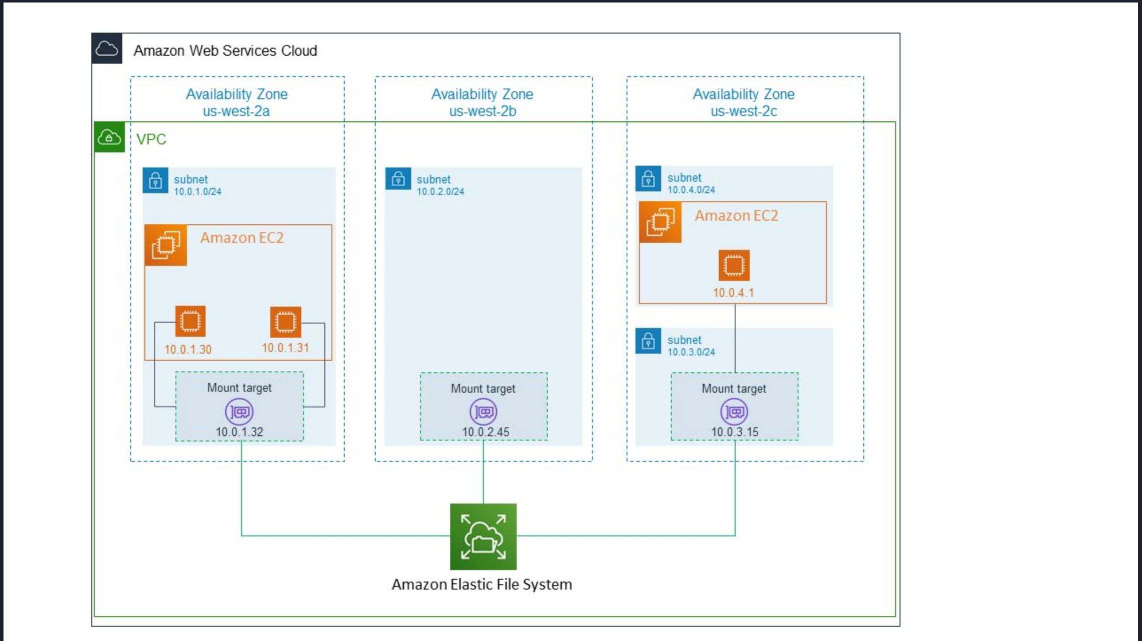
Types of EBS



	Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
API Name	gp2	io1	st1	sc1
Volume Size	1 GiB – 16 TiB	4 GiB – 16 TiB	500 GiB – 16 TiB	500 GiB – 16 TiB
Max. IOPS/Volume	16,000	64,000	500	250
Max. Throughput/Volume	250 MiB/s	1,000 MiB/s	500 MiB/s	250 MiB/s
Max. IOPS/Instance	80,000	80,000	80,000	80,000
Max. Throughput/Instance	1,750 MiB/s	1,750 MiB/s	1,750 MiB/s	1,750 MiB/s
Dominant Performance Attribute	IOPS	IOPS	MiB/s	MiB/s

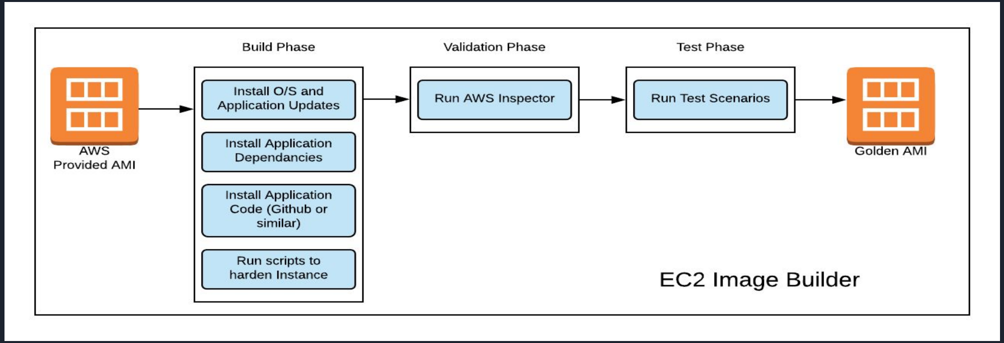
What is EFS?

- Amazon EFS provides scalable file storage for use with Amazon EC2. We can use an EFS file system as a common data source for workloads and applications running on multiple instances.
- EFS works with Linux EC2 instances in multi-AZ



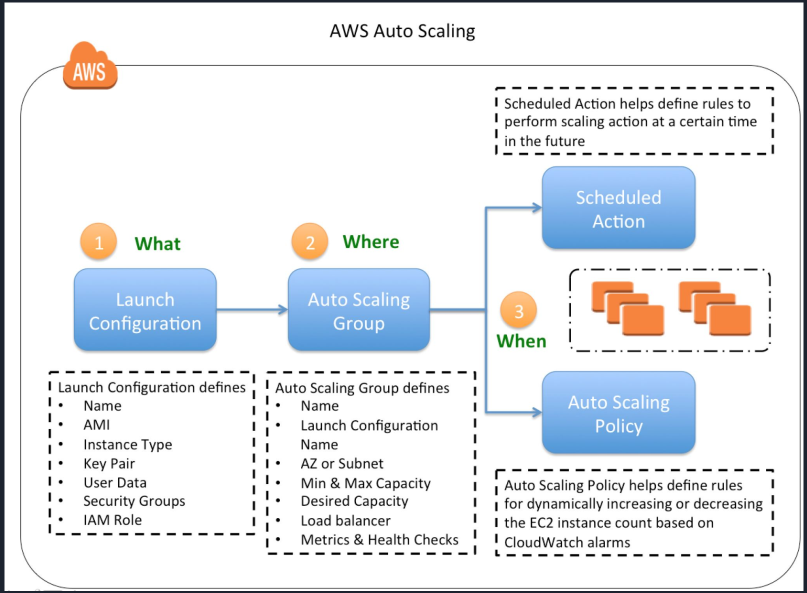
Ec2 AMI Overview

- An Amazon Machine Image (AMI) is a template that contains a software configuration (for example, an operating system, an application server, and applications)
- 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 (Private): you make and maintain them yourself
  - An AWS Marketplace AMI: an AMI someone else made (and potentially sells)
- Launch instances from other AMIs (EC2 Image Builder)
  - Automate the creation, maintain, validate and test EC2 AMIs
  - Can be run on a schedule
  - Free service


Build Phase
Validation Phase
Test Phase
AWS Provided AMI
Install O/S and Application Updates
Install Application Dependancies
Install Application Code (Github or similar)
Run scripts to harden Instance
Run AWS Inspector
Run Test Scenarios
Golden AMI
EC2 Image Builder

Ec2 AMI Overview

- 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)
    - Scale in (remove EC2 instances)
    - Ensure we have a min and a max number of machines running
    - Automatically register new instances to a load balancer
    - Replace unhealthy instances



AWS Auto Scaling

Scheduled Action helps define rules to perform scaling action at a certain time in the future

Launch Configuration defines
- Name
- AMI
- Instance Type
- Key Pair
- User Data
- Security Groups
- IAM Role

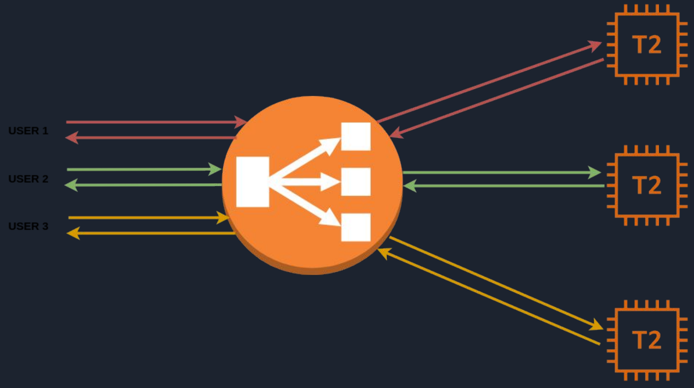
Auto Scaling Group defines
- Name
- Launch Configuration Name
- AZ or Subnet
- Min & Max Capacity
- Desired Capacity
- Load balancer
- Metrics & Health Checks

Auto Scaling Policy helps define rules for dynamically increasing or decreasing the EC2 instance count based on CloudWatch alarms

- Cost Savings: only run at an optimal capacity (principle of the cloud)

What is Elastic Load Balancer?

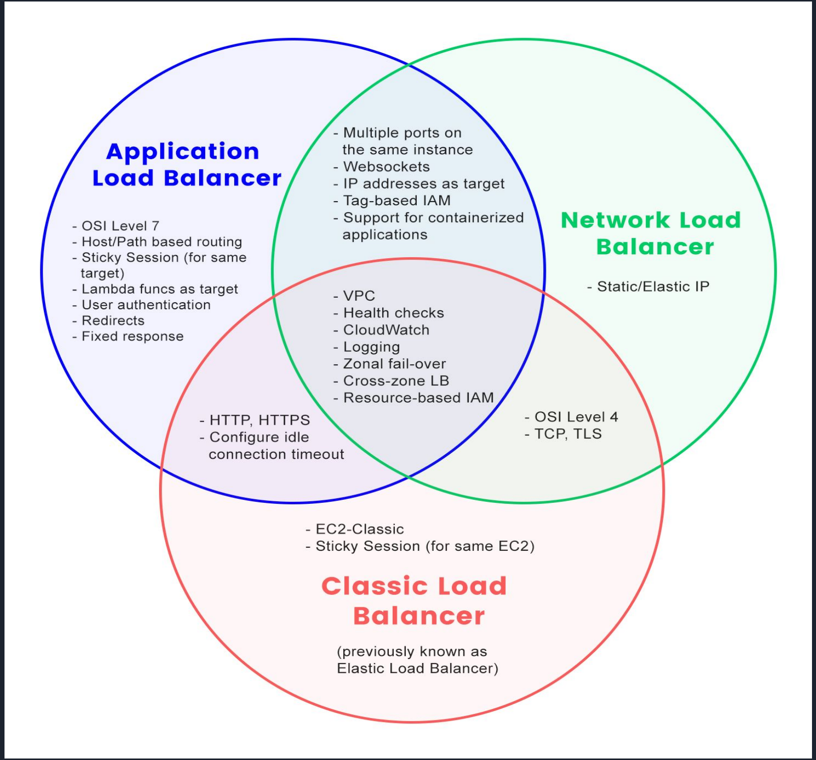
- Load balancers are servers that forward internet traffic to multiple servers (EC2 Instances) downstream
- An ELB (Elastic Load Balancer) is a managed load balancer
- AWS takes care of upgrades, maintenance, high availability
- AWS provides only a few configuration knobs
- It costs less to setup your own load balancer but it will be a lot more effort on your end (maintenance, integrations)



Type of Elastic Load Balancer

	CLB	ALB	NLB
Protocols	TCP, SSL/TLS, HTTP, HTTPS	HTTP, HTTPS	TCP, TLS
Performance (a higher number is slower) the ability to handle more traffic	2	3	1 (fastest)
Host/Path-based routing	No	Yes	No
Sticky Session (for session-based applications)	Yes (redirect to the same machine)	Yes (redirect to the same target)	No
Static/Elastic IP	No	No	Yes
Load balancing to multiple ports on the same instance	No	Yes	Yes
Configurable idle connection timeout	Yes	Yes	No

Type of Elastic Load Balancer



Application Load Balancer

- OSI Level 7
- Host/Path based routing
- Sticky Session (for same target)
- Lambda funcs as target
- User authentication
- Redirects
- Fixed response

- Multiple ports on the same instance
- Websockets
- IP addresses as target
- Tag-based IAM
- Support for containerized applications

Network Load Balancer

- Static/Elastic IP

- VPC
- Health checks
- CloudWatch
- Logging
- Zonal fail-over
- Cross-zone LB
- Resource-based IAM

- HTTP, HTTPS
- Configure idle connection timeout

- OSI Level 4
- TCP, TLS

- EC2-Classic
- Sticky Session (for same EC2)

Classic Load Balancer

(previously known as Elastic Load Balancer)

Thank you!

WOMEN iN BIG DATA
BANGLADESH

aws USER GROUP
BANGLADESH

AWS
User Groups