AWS QUETIONS

https://github.com/vaibhav0342

1. IAM and AWS STS quotas, name requirements, and character limits.

The following quotas are adjustable.

Default quotas for IAM entities

Resource	Default quota	Maximum quota
Customer managed policies in an AWS account	1500	5000
Groups in an AWS account	300	500
Instance profiles in an AWS account	1000	5000
Managed policies attached to an IAM role	10	20
Managed policies attached to an IAM user	10	20
Role trust policy length	2048 characters	4096 characters
Roles in an AWS account	1000	5000
Server certificates stored in an AWS account	20	1000
Virtual MFA devices (assigned or unassigned) in an AWS account	Equal to the user quota for the account	Not applicable

You cannot request an increase for the following quotas.

Quotas for IAM entities

Resource	Quota
Access keys assigned to an IAM user	2
Access keys assigned to the AWS account root user	2
Aliases for an AWS account	1
Groups an IAM user can be a member of	10

Quotas for IAM entities	
Resource	Quota
IAM users in a group	Equal to the user quota for the account
Login profiles for an IAM user	1
Managed policies attached to an IAM group	10
OpenID Connect identity providers per AWS account	100
Permissions boundaries for an IAM user	1
Permissions boundaries for an IAM role	1
MFA devices in use by an IAM user	1
MFA devices in use by the AWS account root user	1
Roles in an instance profile	1
Signing certificates assigned to an IAM user	2
SSH public keys assigned to an IAM user	5
Tags that can be attached to a customer managed policy	50
Tags that can be attached to a SAML identity provider	50
Tags that can be attached to a server certificate	50
Tags that can be attached to a virtual MFA device	50
Tags that can be attached to an instance profile	50
Tags that can be attached to an IAM role	50
Tags that can be attached to an IAM user	50
Tags that can be attached to an Open ID Connect (OIDC) identity provider	50
Users in an AWS account	5000 (If you need to add a large number of users, consider using temporary security credentials.)
Versions of a managed policy that can be stored	5

2. **How much data can be stored in a single bucket?**
   5TB

3. **How many S3 buckets can a user have per account?**
   By default, you can create up to **100 buckets** in each of your AWS accounts. If you need additional buckets, you can increase your account bucket limit to a **maximum of 1,000** buckets by submitting a service limit increase.

4. **Limitation of S3**
   - By default, customers can provision up to 100 buckets per AWS account. However, you can increase your Amazon S3 bucket limit by visiting AWS Service Limits.
   - An object can be 0 bytes to 5TB.
   - The largest object that can be uploaded in a single PUT is 5 gigabytes
   - For objects larger than 100 megabytes, customers should consider using the Multipart Upload capability

5. **Difference between RSA and ED25519**
   RSA is universally supported among SSH clients while EdDSA performs much faster and provides the same level of security with significantly smaller keys.

6. **S3 Encryption Support**
   AWS S3 offers multiple encryption options for stored data.
   First, **server-side encryption (SSE)** can used to secure data-at-rest, which encrypts the incoming object data as it is persisted into the storage layer. It protects user data from prying eyes that have access to the physical media.

   There are three ways to maintain the encryption keys for SSE:
   - Amazon S3-managed Encryption Keys (SSE-S3)
     In this scenario the keys are managed by S3 itself and are only usable for S3 services.

   - KMS-managed Encryption Keys (SSE-KMS)
     Here the keys are managed by a **key management system (KMS)**, which is a separate AWS service (part of the AWS IAM service). These keys can be used for multiple AWS services, including S3. Note that each object is encrypted using a dedicated key, and the KMS key is used to secure the per-object keys.

   - Customer-provided Encryption Keys (SSE-C)
     The last option is for users to provide a key when writing and reading S3 objects. These keys have no relation to anything else in S3.

   Of note is that SSE is only protecting the data at rest, that is, how the data is delivered is a separate concern.

This leads to the second encryption option for S3, which is **client-side encryption (CSE)**. Here the client is tasked to encrypt the data before it is sent to S3 and handles the security of data-in-transit.

There are two ways of maintaining encryption keys for CSE:

- KMS-managed Encryption Keys (CSE-KMS)
  This is the same as for SSE, that is, the KMS in the IAM service is used to manage shared keys. The client can use the key ID (the ARN) to refer to a key, which is then accessed by the client for encryption and decryption purposes. Note that each object is encrypted using a dedicated key, and the KMS key is used to secure the per-object keys.

- Client-side Master Keys (CSE-C)
  Another option is to provide a local, client-side only key that is used to encrypt and decrypt locally. These keys are having no relation to anything else in AWS.

7. SSH

- **Generate ssh key without any arguments**
  # ssh-keygen

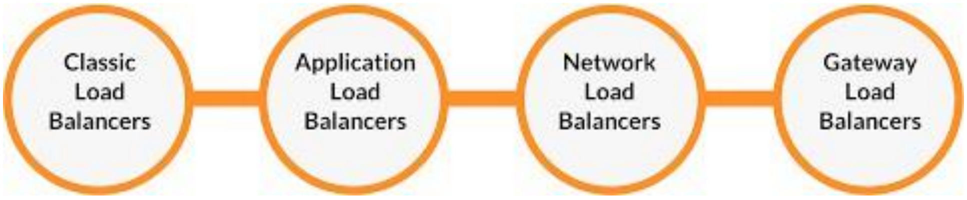
- **Define Key Type**
  By default, ssh-keygen will create RSA type key
  You can create key with dsa, ecdsa, ed25519, or rsa type
  Use -t <key> argument to define the type of the key
  # ssh-keygen -t ed25519

- **Define Bit size**
  By default, ssh-keygen generates SSH key with 2048bit size. You can also specify the number of bits to be used for the keys by using -b <bit_size>
  # ssh-keygen -b 4096

8. Difference Between Network Load Balancer and Application Load Balancer

Elastic load Balancer?
Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets and virtual appliances in one or more Availability Zones (AZs)

Types of load balancers

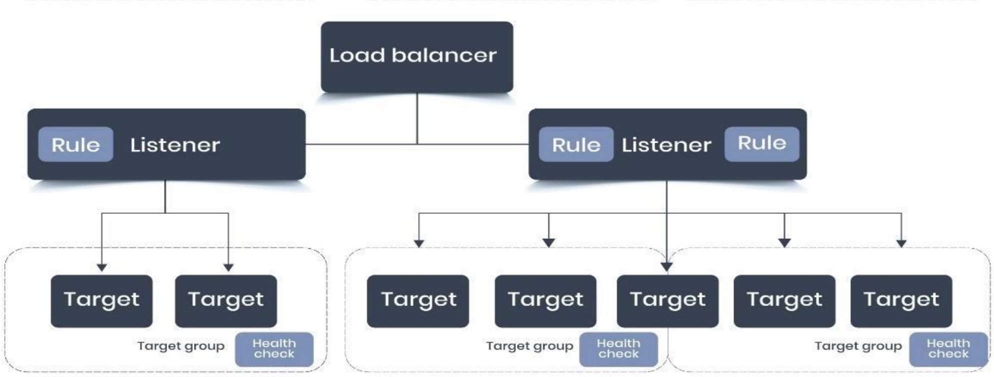


- **Application Load Balancer**
  Application Load Balancer operates on the 7th layer of the OSI Model as the name suggests. It has the ability to examine the application-level content and

route the traffic based on this acquired information. Application-level content includes packet details, HTTP and HTTPS details. This makes the routing easier, faster and much more efficient. It’s one of the most widely used ELB.

Working Of Application Load Balancer

Application Load Balancer consists of listeners and rules. When a client makes the request, the listener acknowledges it. The rules are guidelines that govern the routing of each client request once it’s heard by the listener. The rules consist of three components – Target group, Priority and Conditions. Target Groups consists of registered targets (servers where the traffic is to be routed). Each target group routes requests to one or more registered targets, such as EC2 instances, using the protocol and port number that you specify. So basically, when the listener gets the request, it goes through priority order to determine which rule to apply, analyses the rules and based on condition, decides which target group gets the request.



➢ **Network Load Balancer**

Network Load Balancers use variables such as destination ports and IP addresses to distribute traffic. They function on OSI Layer 4, so they are not intended to be context-aware or to consider cues at the application layer such as cookie data, content type, user location, custom headers, or application behaviour. Network Load Balancers consider only the network-layer information contained inside the packets they direct.

	Network Load Balancer	**Application Load Balancer**
How it works	Directs traffic based on IP addresses and transport layer ports, without visibility into the content of protocol data units sent or received,	Examines incoming traffic from internet facing applications and direct request based on application layer-based protocol request header data.
Where it operates in OSI model	Layer 4, the transport layer.	Layer 7, the application layer.
Supported protocols	TCP, UDP, TLS	HTTP, HTTPS, SMTP

Common algorithms	Round Robin, Weighted Round Robin, least connection, weighted least connection etc.	Round Robin, least outstanding request.
Benefits	Doesn’t examine the content of higher layer protocols (e.g., HTTPS requests/responses or TLS encryption) so it can be faster than application load balancer.	Handles more complex routing decisions and provides greater efficiency because it identifies the request load.

- **Gateway Load Balancer**

  Gateway Load Balancers enable you to deploy, scale, and manage virtual appliances, such as firewalls, intrusion detection and prevention systems, and deep packet inspection systems. It combines a transparent network gateway (that is, a single entry and exit point for all traffic) and distributes traffic while scaling your virtual appliances with the demand.

  A Gateway Load Balancer operates at the third layer of the Open Systems Interconnection (OSI) model, the network layer. It listens for all IP packets across all ports and forwards traffic to the target group that's specified in the listener rule. It maintains stickiness of flows to a specific target appliance using 5-tuple (for TCP/UDP flows) or 3-tuple (for non-TCP/UDP flows). The Gateway Load Balancer and its registered virtual appliance instances exchange application traffic using the GENEVE protocol on port 6081.

- **Classical Load Balancer**

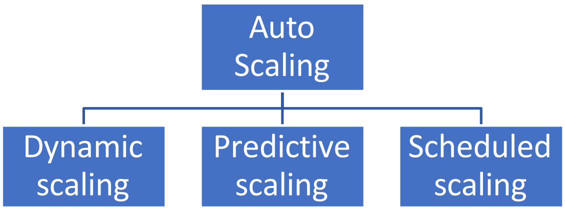
  Classic Load Balancer provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level. Classic Load Balancer is intended for applications that are built within the EC2-Classic network. We recommend Application Load Balancer for Layer 7 traffic and Network Load Balancer for Layer 4 traffic when using Virtual Private Cloud (VPC).

9. What is listener?

A listener is a process that checks for connection requests, using the protocol and port that you configure.

10. What is autoscaling and its type.

Auto-scaling is a way to automatically scale up or down the number of compute resources that are being allocated to your application based on its needs at any given time.



- **Dynamic scaling**
  When you configure dynamic scaling, you define how to scale the capacity of your Auto Scaling group in response to changing demand.
  For example, let's say that you have a web application that currently runs on two instances, and you want the CPU utilization of the Auto Scaling group to stay at around 50 percent when the load on the application changes. This gives you extra capacity to handle traffic spikes without maintaining an excessive number of idle resources.

- **Predictive scaling**
  Use predictive scaling to increase the number of EC2 instances in your Auto Scaling group in advance of daily and weekly patterns in traffic flows.
  In general, if you have regular patterns of traffic increases and applications that take a long time to initialize, you should consider using predictive scaling. Predictive scaling can help you scale faster by launching capacity in advance of forecasted load, compared to using only dynamic scaling, which is reactive in nature. Predictive scaling can also potentially save you money on your EC2 bill by helping you avoid the need to overprovision capacity.

- **Scheduled scaling**
  Scheduled scaling helps you to set up your own scaling schedule according to predictable load changes. For example, let's say that every week the traffic to your web application starts to increase on Wednesday, remains high on Thursday, and starts to decrease on Friday. You can configure a schedule for Amazon EC2 Auto Scaling to increase capacity on Wednesday and decrease capacity on Friday

11. Policies and permissions in IAM

You manage access in AWS by creating policies and attaching them to IAM identities (users, groups of users, or roles) or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions.

Policy types

1) **Identity-based policies** – Attach managed and inline policies to IAM identities (users, groups to which users belong, or roles). Identity-based policies grant permissions to an identity.
2) **Resource-based policies** – Attach inline policies to resources. The most common examples of resource-based policies are Amazon S3 bucket policies and IAM role trust policies. Resource-based policies grant permissions to the principal that is specified in the policy. Principals can be in the same account as the resource or in other accounts.

3) **Permissions boundaries** – Use a managed policy as the permissions boundary for an IAM entity (user or role). That policy defines the maximum permissions that the identity-based policies can grant to an entity, but does not grant permissions. Permissions boundaries do not define the maximum permissions that a resource-based policy can grant to an entity.

4) **Organizations SCPs** – Use an AWS Organizations service control policy (SCP) to define the maximum permissions for account members of an organization or organizational unit (OU). SCPs limit permissions that identity-based policies or resource-based policies grant to entities (users or roles) within the account, but do not grant permissions.

5) **Access control lists (ACLs)** – Use ACLs to control which principals in other accounts can access the resource to which the ACL is attached. ACLs are similar to resource-based policies, although they are the only policy type that does not use the JSON policy document structure. ACLs are cross-account permissions policies that grant permissions to the specified principal. ACLs cannot grant permissions to entities within the same account.

6) **Session policies** – Pass advanced session policies when you use the AWS CLI or AWS API to assume a role or a federated user. Session policies limit the permissions that the role or user's identity-based policies grant to the session. Session policies limit permissions for a created session, but do not grant permissions. For more information, see Session Policies.

Identity-based policies

Identity-based policies are JSON permissions policy documents that control what actions an identity (users, groups of users, and roles) can perform, on which resources, and under what conditions. Identity-based policies can be further categorized:

- **Managed policies** – Standalone identity-based policies that you can attach to multiple users, groups, and roles in your AWS account. There are two types of managed policies:
  - AWS managed policies – Managed policies that are created and managed by AWS.
  - Customer managed policies – Managed policies that you create and manage in your AWS account. Customer managed policies provide more precise control over your policies than AWS managed policies.

- **Inline policies** – Policies that you add directly to a single user, group, or role. Inline policies maintain a strict one-to-one relationship between a policy and an identity. They are deleted when you delete the identity**.**

NOTE - *The different types of policies are for different use cases. In most cases, we recommend that you use managed policies instead of inline policies.*

Inline policies are useful if you want to maintain a strict one-to-one relationship between a policy and the identity to which it is applied. For example, you want to be sure that the permissions in a policy are not inadvertently assigned to an identity other than the one they're intended for. When you use an inline policy, the permissions in the policy cannot be inadvertently attached to the wrong identity. In addition, when you use the AWS Management Console to delete that identity, the policies embedded in the identity are deleted as well. That's because they are part of the principal entity.

12. **What is trust relationship?**

Roles are used to grant specific privileges to specific actors for a set duration of time. So, a role needs two things: permission policies (what resources can be accessed and what actions can be taken) and a trust policy (what entities can assume the role).

I want to be able to download some configuration file from s3 bucket into my web application, the web application runs on ec2 instance and the s3 bucket name is "configuration-for-app"

I'm creating a role named "my-app-role" which contains several policies, one of them is s3 policy that can access my s3 amazon resource "configuration-for-app" and has explicit permission to get it only (not delete it, not changing it - just get it). **Since the app runs on ec2 - the trusted relations in these requirements between these services would be <ec2> -> <s3>**, my application that runs on ec2 can assume that role (my-app-role) and accessing (with the correct policy in it) to s3 and get the configuration file.

13. **What is tenancy?**

Tenancy defines how EC2 instances are distributed across physical hardware and affects pricing. There are three tenancy options available:

- Shared (`default`) — Multiple Amazon Web Services accounts may share the same physical hardware.
- Dedicated Instance (`dedicated`) — Your instance runs on single-tenant hardware.
- Dedicated Host (`host`) — Your instance runs on a physical server with EC2 instance capacity fully dedicated to your use, an isolated server with configurations that you can control.

14. **Difference between IAAS, PAAS and SAAS**

IaaS	PaaS	SaaS
Infrastructure as a service	Platform as a service	Software as a service
A service model in cloud computing that provides virtualized computing resources over the internet.	A cloud computing model that delivers tools necessary for application development over the internet.	A service model in cloud computing that hosts software and makes them available for clients over the internet.

Provides access to resources such as virtual machines, virtual storage etc.	Provides runtime environments, development and deployment tools for applications.	Provides software as service to the users.
Use by network architects.	Use by developers,	Use by end users.

15. What is internet gateway?

An internet gateway enables resources (like EC2 instances) in your public subnets to connect to the internet if the resource has a public IPv4 address or an IPv6 address.

16. What is NAT gateway?

A NAT gateway in a device forwards the traffic from instances present in the private subnet to the internet/AWS services, and sends back the response from the server back to the instance. When the traffic moves to the internet, an IPV4 address gets replaced with the NAT's device address. Once the response is obtained, it has to be sent to the instance, and in this case, the NAT device translates the address back to the IPV4 and it is given to the IPV4 address.

There are two kinds of NAT devices which AWS offers- A NAT gateway and a NAT instance. AWS recommends the usage of NAT gateways since it helps provide high availability and a better bandwidth in comparison to NAT instance.

17. Difference between Security Groups and Network Access Control List

1. Security Group:

Security group like a virtual firewall. It has inbound and outbound security rules in which all inbound traffic is blocked by default in private on AWS EC2. It does not allow particular protocol no one will able to access our instances using this protocol you can stop traffic by using that rule by default everything that is denied. There are various multiple security groups on EC2 instances. We cannot block a specific IP address using that security group but using the network access list. In which we edit any rule a security group with faster effect.

2. Network Access Control List (Network ACL):

Network ACL is a modifiable default network. It allows all the inbound or outbound IPv4 traffic and here we create a type of custom network all or each custom network ACL denies all inbound and outbound traffic. This network is the stateless and separate inbound and outbound rule with a default limit of 20 for both rules and starting with the lowest numbered rule. In which all subnet in VPC must be combined with network ACL one subnet -one network ACL at a time. It supports rules and deny rules and operate the subnet level.

Security Group	Network Access Control List
In security group, we operate at instance level.	In network ACL, we operate sub net level.
It supports only allow rules.	It supports allow rules and deny rules.

It is stateful, when we create an inbound or an outbound rule.	It is stateless, it returns traffic must be allowed explicitly.
We cannot block specific IP address using SGs.	We can block specific IP Address using NACL.
All rules are evaluated before deciding to permit traffic.	Rules are processed in number order when deciding whether allow traffic.
It starts with instance launch configuration.	In which we assigned to subnet for all instance.
It applies when someone specifies security group when launching the instance and it associates with security group.	They do not depend on user it automatically applies all instances with subnet.

18. What is Placement Groups?

Placement groups are a way of a logical grouping of interdependent instances together in a selected region, or in other words as the name implies, a placement group is just a group. AWS instances that exist within a common availability zone can be grouped into a placement group. Where group members are able to communicate with one another in a way that provides low latency and high throughput.

- Why use Placement Group?
  Placement groups help us to launch a bunch of EC2 instances close to each other physically within the same AZ. Being close physically and within the same AZ helps it take advantage of high-speed connectivity to provide low latency, high throughput access.
  This can really work well for applications exchanging a lot of data and can provide high performance with collocation.
- When to use a Placement Group?
  When applications reaching oversubscriptions or when there is a high latency etc.
- Types of Placement Groups?
  we can create a placement group using one of the following placement strategies:
  - I. **Cluster Placement Group:** Packs instances close together inside an Availability Zone. This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.
  - II. **Partition Placement Group:** Spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.
  - III. **Spread Placement Group:** Strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

19. What is AWS certification?

AWS stands for Amazon Web Services and it's a cloud base services to host IT systems, Infra etc. AWS service is offered by Amazon.

Amazon has AWS certification programs at various level i.e., Specialty, Professionals and Associates. The below are the available certifications by AWS:

- **Associate**
  1. Certified Solutions Architect Associate
  2. Certified Developer Associate
  3. Certified Sysops Administrator Associate
- **Professional**
  1. Certified Solutions Architect Professional
  2. Devops Professional
- **Specialty**
  1. Security
  2. Advanced Networking
  3. BiG Data

20. What is health check?

Health checks are a way of asking a service on a particular server whether or not it is capable of performing work successfully. Load balancers ask each server this question periodically to determine which servers it is safe to direct traffic to.

21. What is Hypervisor?

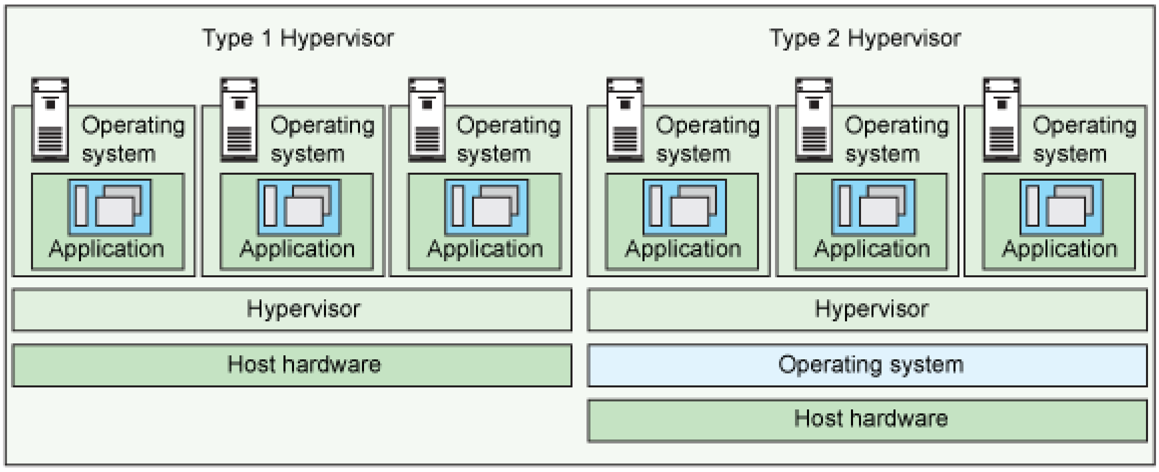
A hypervisor, also known as a virtual machine monitor or VMM, is software that creates and runs virtual machines (VMs). A hypervisor allows one host computer to support multiple guest VMs by virtually sharing its resources, such as memory and processing.

Virtualization

Virtualization is the process of running a virtual instance of a computer system in a layer abstracted from the actual hardware. Most commonly, it refers to running multiple operating systems on a computer system simultaneously.

There are two types of hypervisors:

- Type 1 hypervisor: hypervisors run directly on the system hardware – A "bare metal" embedded hypervisor,
- Type 2 hypervisor: hypervisors run on a host operating system that provides virtualization services, such as I/O device support and memory management.



The most commonly deployed type of hypervisor is the type 1 or bare-metal hypervisor, where virtualization software is installed directly on the hardware where the operating system is normally installed. Because bare-metal hypervisors are isolated from the attack-prone operating system, they are extremely secure. In addition, they generally perform better and more efficiently than hosted hypervisors. For these reasons, most enterprise companies choose bare-metal hypervisors for data center computing needs.

While bare-metal hypervisors run directly on the computing hardware, hosted hypervisors run on top of the operating system (OS) of the host machine. Although hosted hypervisors run within the OS, additional (and different) operating systems can be installed on top of the hypervisor. The downside of hosted hypervisors is that latency is higher than bare-metal hypervisors. This is because communication between the hardware and the hypervisor must pass through the extra layer of the OS. Hosted hypervisors are sometimes known as client hypervisors because they are most often used with end users and software testing, where higher latency is less of a concern.

Type 1 hypervisors:

1. VMware ESX and ESXi
2. Microsoft Hyper-V
3. Citrix XenServer
4. Oracle VM

Type 2 hypervisor

1. VMware Workstation/Fusion/Player
2. VMware Server
3. Microsoft Virtual PC
4. Oracle VM VirtualBox
5. Red Hat Enterprise Virtualization

22. What are global, regional and AZ resources services?

AWS Global, Regional, AZ resource Availability

AWS provides a lot of services and these services are either Global, Regional or specific to the Availability Zone and cannot be accessed outside. Most of the AWS managed services are regional based services (except for IAM, Route53, CloudFront, WAF etc).

23. Global vs Regional vs AZ Resource locations

- **IAM**
    - **Users, Groups, Roles, Accounts** – Global
        - Same AWS accounts, users, groups and roles can be used in all regions
    - **Key Pairs** – Regional
        - Amazon EC2 created key pairs are specific to the region
        - RSA key pair can be created and uploaded that can be used in all regions

- **Virtual Private Cloud**
    - VPC – Regional
        - VPC are created within a region
    - Subnet – Availability Zone
        - Subnet can span only a single Availability Zone
    - Security groups – Regional
        - A security group is tied to a region and can be assigned only to instances in the same region.
    - VPC Endpoints – Regional
        - You cannot create an endpoint between a VPC and an AWS service in a different region.
    - VPC Peering
        - VPC Peering can be performed across VPC in the same account or different AWS accounts. VPC Peering can now span inter-region
    - Elastic IP Address – Regional
        - Elastic IP address created within the region can be assigned to instances within the region only

- **EC2**
    - Resource Identifiers – Regional
        - Each resource identifier, such as an AMI ID, instance ID, EBS volume ID, or EBS snapshot ID, is tied to its region and can be used only in the region where you created the resource.
    - Instances – Availability Zone
        - An instance is tied to the Availability Zones in which you launched it. However, note that its instance ID is tied to the region.
    - EBS Volumes – Availability Zone
        - Amazon EBS volume is tied to its Availability Zone and can be attached only to instances in the same Availability Zone.
    - EBS Snapshot – Regional
        - An EBS snapshot is tied to its region and can only be used to create volumes in the same region and has to be copied from One region to other if needed

  - AMIs – Regional
    - AMI provides templates to launch EC2 instances
    - AMI is tied to the Region where its files are located with Amazon S3. For using AMI in different regions, the AMI can be copied to other regions
  - Auto Scaling – Regional
    - Auto Scaling spans across multiple Availability Zones within the same region but cannot span across regions
  - Elastic Load Balancer – Regional
    - Elastic Load Balancer distributes traffic across instances in multiple Availability Zones in the same region
  - Cluster Placement Groups – Availability Zone
    - Cluster Placement groups can be span across Instances within the same Availability Zones

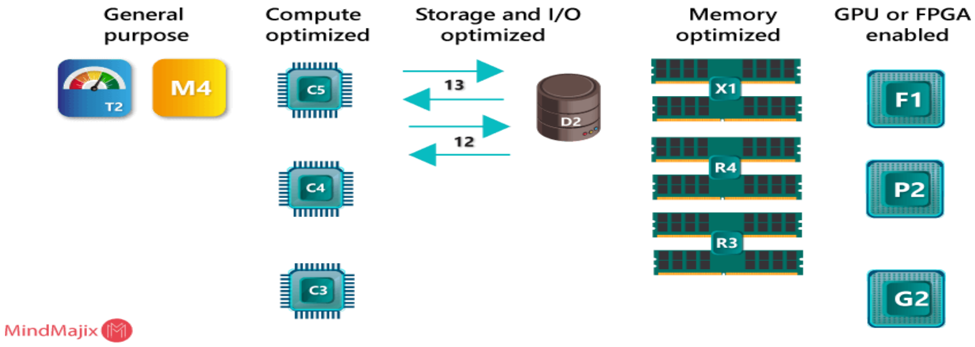
- **S3** – Global but Data is Regional
  - S3 buckets are created within the selected region
  - Objects stored are replicated across Availability Zones to provide high durability but are not cross region replicated unless done explicitly
  -
- **Route53** – Global
  - Route53 services are offered at AWS edge locations and are global

- **DynamoDb** – Regional
  - All data objects are stored within the same region and replicated across multiple Availability Zones in the same region
  - Data objects can be explicitly replicated across regions using cross-region replication
- **WAF** – Global
  - Web Application Firewall (WAF) services protects web applications from common web exploits are offered at AWS edge locations and are global

- **CloudFront** – Global
  - CloudFront is the global content delivery network (CDN) services are offered at AWS edge locations

- **Storage Gateway** – Regional
  - AWS Storage Gateway stores volume, snapshot, and tape data in the AWS region in which the gateway is activated

24. Type of instances?



- **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.

- **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 modelling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.

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

- **Accelerated Computing**
  Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating-point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs.

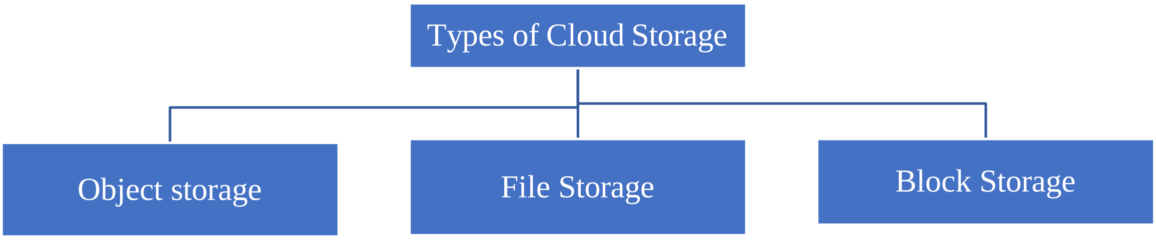
- **Storage Optimized**
  Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latencies, random I/O operations per second (IOPS) to applications.

25. What is an Egress only Internet Gateway?

- An egress-only internet gateway is a horizontally scaled, redundant, and highly available VPC component.
- It allows outbound communication over IPv6 from instances in your VPC to the internet, and prevents the internet from initiating an IPv6 connection with your instances.

- IPv6 addresses are globally unique, and are therefore public by default.
- If you want your instance to be able to access the internet, but you want to prevent resources on the internet from initiating communication with your instance, you can use an egress-only internet gateway.
- An egress-only internet gateway is stateful: it forwards traffic from the instances in the subnet to the internet or other AWS services, and then sends the response back to the instances.

26. Types of Cloud Storage



27. AWS storage services

1) **Object, file, and block storage**
   - Amazon Simple Storage Service (**S3**)
   - Amazon Elastic File System (**EFS**)
   - Amazon FSx
   - Amazon Elastic block Store (**EBS**)

2) **Data Migration**
   - AWS DataSync
   - AWS Snow Family

3) **Hybrid cloud storage and edge computing**
   - AWS Storage Gateway
   - AWS Snow Family

4) **Managed file transfer**
   - AWS Transfer Family

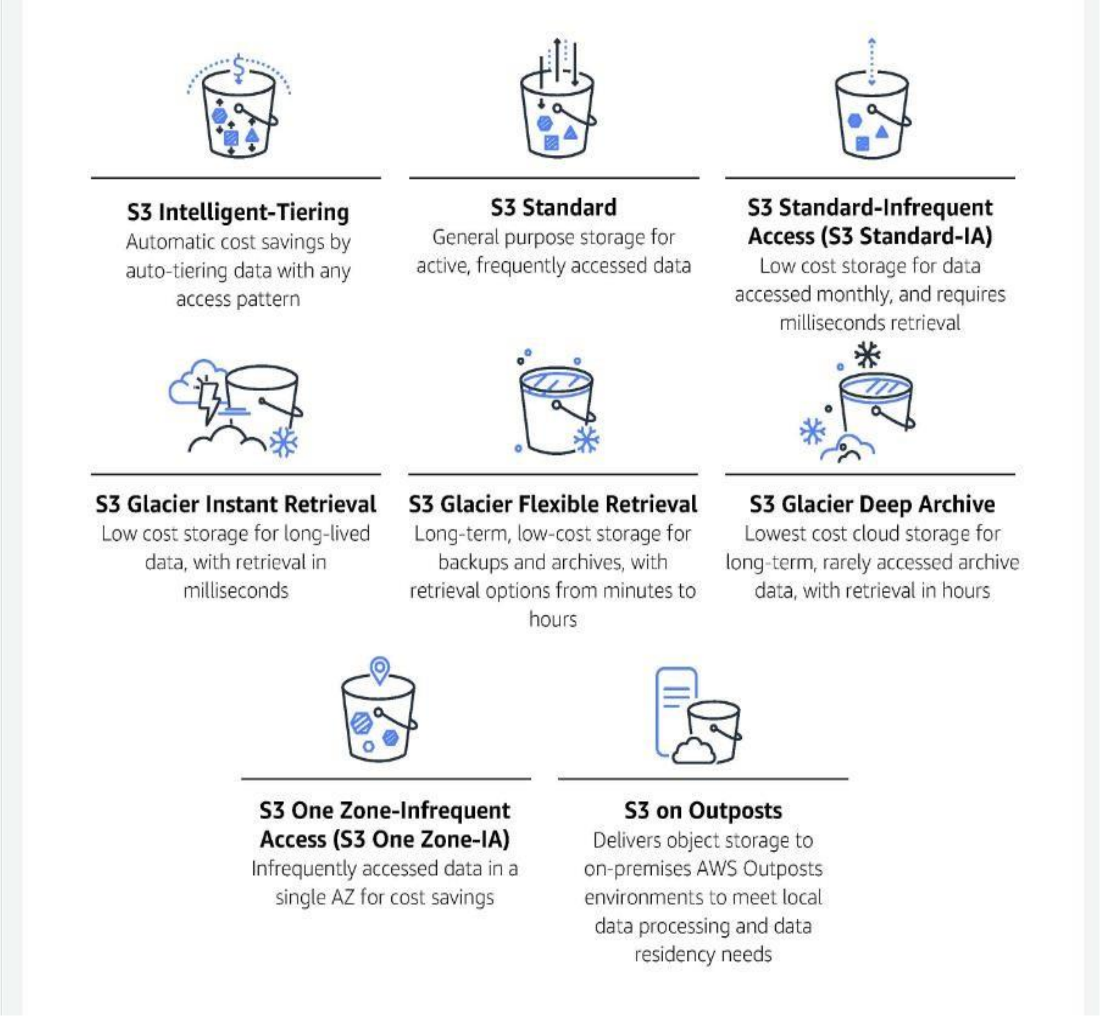
5) **Disaster recovery and backup**
   - AWS Elastic Disaster Recovery **(EDR)**
   - AWS Backup

28. How many subnets can be created per Amazon VPC?

Currently you can create 200 subnets per VPC.

29. What are S3 storage classes?

Amazon S3 comes in eight 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

- **S3 Intelligent-Tiering**

It is most suitable for data with access needs that are either changing or unknown. S3 Intelligent-Tiering has four different access tiers: Frequent Access, Infrequent Access (IA), Archive and Deep Archive. Data is automatically moved to the most inexpensive storage tier according to customer access patterns.

- **S3 Standard**

It is suitable for frequently accessed data that needs to be delivered with low latency and high throughput. S3 Standard targets applications, dynamic websites, content distribution and big data workloads.

- **S3 Standard-IA**

It offers a lower storage price for data that is needed less often but that must be quickly accessible. This tier can be used for backups and long-term data storage.

- **S3 Glacier Instant Retrieval**

Amazon S3 Glacier Instant Retrieval is an archive storage class that delivers the lowest-cost storage for long-lived data that is rarely accessed and requires retrieval in milliseconds. With S3 Glacier Instant Retrieval, you can save up to 68% on storage costs compared to using the S3 Standard-Infrequent Access (S3 Standard-IA) storage class, when your data is accessed once per quarter.

- **S3 Glacier Flexible Retrieval**

S3 Glacier Flexible Retrieval delivers low-cost storage, up to 10% lower cost (than S3 Glacier Instant Retrieval), for archive data that is accessed 1—2 times per year and is retrieved asynchronously.
S3 Glacier Flexible Retrieval delivers the most flexible retrieval options that balance cost with access times ranging from minutes to hours and with free bulk retrievals.

- **S3 Glacier Deep Archive**

S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class and supports long-term retention and digital preservation for data that may be accessed once or twice in a year. It is designed for customers—particularly those in highly-regulated industries, such as financial services, healthcare, and public sectors—that retain data sets for 7—10 years or longer to meet regulatory compliance requirements

- **S3 One Zone-IA**

S3 One Zone-IA is for data that is accessed less frequently, but requires rapid access when needed. Unlike other S3 Storage Classes which store data in a minimum of three Availability Zones (AZs), S3 One Zone-IA stores data in a single AZ and costs 20% less than S3 Standard-IA.
Use of S3 One Zone-IA is indicated for infrequently accessed data without high resilience or availability needs, data that is able to be recreated and backing up on-premises data.

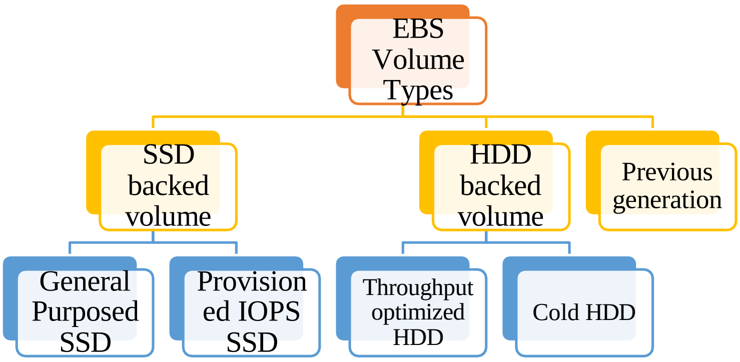
- **S3 Outposts**

It adds S3 object storage features and APIs to an on-premises AWS Outposts environment. S3 Outposts is best used when performance needs call for data to be stored near on-premises applications or to satisfy specific data residency requirements

30. What is EBS storage?

Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with EC2 instances. EBS volumes behave like raw, unformatted block devices.

- EBS is the logical volumes to use with the EC2 instances.
- This type of storage is used, when the data needs to be accessed quickly and its requirement is for a long duration of time.
- You can create a file system on top of these volumes.
- A lifetime of the EBS is not dependent on the EC2 instance.
- Volume and instance must be in the same Availability Zone.
- A volume can be attached to only one instance at a time.
- It can be detached and attached between the instances in the same Availability Zone.
- But an Instance can be attached N number of volumes.



Solid state drives (SSD)

The SSD-backed volumes provided by Amazon EBS fall into these categories:

- General Purpose SSD — Provides a balance of price and performance. We recommend these volumes for most workloads.
- Provisioned IOPS SSD — Provides high performance for mission-critical, low-latency, or high-throughput workloads.

Hard disk drives (HDD)

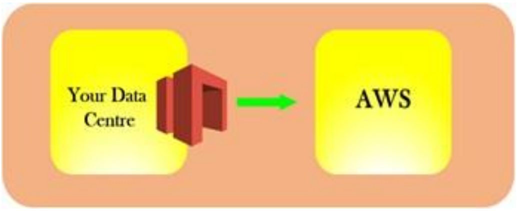
The HDD-backed volumes provided by Amazon EBS fall into these categories:

- Throughput Optimized HDD — A low-cost HDD designed for frequently accessed, throughput-intensive workloads.
- Cold HDD — The lowest-cost HDD design for less frequently accessed workloads.

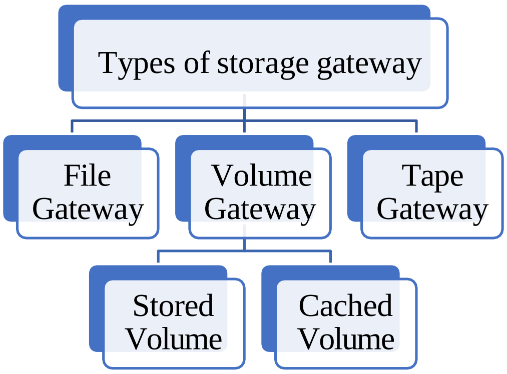
31. What is AWS Storage Gateway (Amazon Web Services Storage Gateway)

AWS Storage Gateway in an Amazon Web Services tool that connects on-premises software resources with storage in the AWS public cloud. The service securely transfers application data between on-premises software and cloud-based storage to improve application scalability and help an enterprise save costs.

- Storage Gateway service allows you to securely store the data in AWS cloud for the scalable and cost-effective storage.
- Storage Gateway is a virtual appliance which is installed in a hypervisor running in a Data center used to replicate the information to the AWS particularly S3.
- Amazon Storage Gateway's virtual appliance is available for download as a virtual machine (VM) image which you can install on a host in your data center.
- Storage Gateway supports either Vmware EXI or Microsoft Hyper-V.
- Once you have installed the storage gateway, link it with your AWS account through the activation process, and then you can use the AWS Management Console to create the storage gateway option.



There are three types of Storage Gateways



- **File Gateway**
  - It is using the technique NFS.
  - It is used to store the flat files in S3 such as word files, pdf files, pictures, videos, etc.
  - It is used to store the files to S3 directly.
  - Files are stored as objects in S3 buckets, and they are accessed through a Network File System (NFS) mount point.
  - Ownership, permissions, and timestamps are durably stored in S3 in the user metadata of the object associated with the file.
  - Once the objects are transferred to the S3, they can be used as the native S3 objects, and bucket policies such as versioning, lifecycle management, and cross-region replication can be directly applied to the objects stored in your bucket.
  - Storage Gateway is a virtual machine running on-premises.
  - Storage Gateway is mainly connected to aws through the internet.
  - It can use Direct Connect. Direct Connect is a direct connection line between the Data center and aws.
  - It can also use an Amazon VPC (Virtual Private Cloud) to connect a storage gateway to aws. VPC is a virtual data center. It represents that the Application server and storage gateway do not need to be on-premises. In Amazon VPC, storage gateway sits inside the VPC, and then storage gateway sends the information to S3.

- **Volume Gateway**
  - Volume Gateway is an interface that presents your applications with disk volumes using the Iscsi block protocol. The iSCSI block protocol is block-based storage that can store an operating system, applications and also can run the SQL Server, database.

- Data written to the hard disk can be asynchronously backed up as point-in-time snapshots in your hard disks and stored in the cloud as EBS snapshots where EBS (Elastic Block Store) is a virtual hard disk which is attached to the EC2 instance. In short, we can say that the volume gateway takes the virtual hard disks that you back them up to the aws.
- Snapshots are incremental backups so that the changes made in the last snapshot are backed up. All snapshot storage is also compressed to minimize your storage charges.

Gateway-cached volumes allow users to store primary data in Amazon Simple Storage Service (S3) while keeping frequently accessed data on-premises. This configuration allows a user to store up to 32 volumes with 32 TB of data per volume. Gateway-cached volumes also allow a user to take snapshots of volume data for protection.

Gateway-stored volumes maintain access to the entire data set. This configuration stores data locally and backs up snapshots to S3 for disaster recovery. A user is limited to 32 volumes with this configuration and a maximum of 16 TB per volume.

➢ **Tape Gateway**

- Tape Gateway is mainly used for taking backups.
- It uses a Tape Gateway Library interface.
- Tape Gateway offers a durable, cost-effective solution to archive your data in AWS cloud.
- The VTL interface provides a tape-based backup application infrastructure to store data on virtual tape cartridges that you create on your tape Gateway.
- It is supported by NetBackup, Backup Exec, Veeam, etc. Instead of using physical tape, they are using virtual tape, and these virtual tapes are further stored in Amazon S3.
- Servers are connected to the Backup Application, and the Backup Application can be NetBackup, Backup Exec, Veeam, etc.
- Backup Application is connected to the Storage Gateway over the iSCSI connection.
- Virtual Gateway is represented as a virtual appliance connected over iSCSI to the Backup application.
- Virtual tapes are uploaded to an Amazon S3.
- Now, we have a Lifecycle Management policy where we can archive to the virtual tape shelf in Amazon Glacier.

Important points to remember:

✓ File Gateway is used for object-based storage in which all the flat files such as word files, pdf files, etc, are stored directly on S3.

✓ Volume Gateway is used for block-based storage, and it is using an iSCSI protocol.

- ✓ Stored Volume is a volume gateway used to store the entire dataset on site and backed up to S3.
- ✓ Cached volume is a volume gateway used to store the entire dataset in a cloud (Amazon S3) and only the most frequently accessed data is kept on site.
- ✓ Tape Gateway is used for backup and uses popular backup applications such as NetBackup, Backup Exec, Veeam, etc.

32. What is Target Group?

A target group is a collection of endpoints or servers which the load balancer can route traffic to. Usually, the load balancer will check a specific URL on each target node within the group to make sure it's healthy before routing traffic to it, but this can be configured to always route traffic even when the target endpoint is unhealthy.

33. What is VPC endpoint?

VPC endpoint enables creation of a private connection between VPC to supported AWS services and VPC endpoint services powered by Private Link using its private IP address. Traffic between VPC and AWS service does not leave the Amazon network.

There are two types of VPC endpoints:

Interface endpoint is an elastic network interface (ENI) with a private IP address from the IP address range of user's subnet that serves as an entry point for traffic destined to a supported service. It enables you to privately access services by using private IP addresses.

Gateway endpoint is a gateway that you specify as a target for a route in your route table for traffic destined to a supported AWS service. Currently supports S3 and DynamoDB services.

VPC Endpoints Limitations

- ➢ VPC endpoints support IPv4 traffic only.
- ➢ Endpoints are supported within the same Region only. You cannot create an endpoint between a VPC and a service in a different Region.
- ➢ Endpoints cannot transfer an endpoint from one VPC to another, or from one service to another.

34. What are AWS IAM?

AWS Identity and Access Management (IAM) is a web service that helps you securely control access to AWS resources for your users. You use IAM to control who can use your AWS resources (authentication) and how they can use resources (authorization).

35. What are policies?

A policy is an object in AWS that, when associated with an entity or resource, defines their permissions.

36. **IAM roles**

An IAM role is an IAM identity that you can create in your account that has specific permissions. An IAM role is similar to an IAM user, in that it is an AWS identity with permission policies that determine what the identity can and cannot do in AWS.

37. **What is the difference between terminating and stopping an EC2 instance?**

Background Information

Amazon supports the ability to terminate or stop a running instance. The ability to stop a running instance is only supported by instances that were launched with an EBS-based AMI.

- **Terminate Instance**

  When you terminate an EC2 instance, the instance will be shut down and the virtual machine that was provisioned for you will be permanently taken away and you will no longer be charged for instance usage. Any data that was stored locally on the instance will be lost. Any attached EBS volumes will be detached and deleted. However, if you attach an EBS Snapshot to an instance at boot time, the default option in the Dashboard is to delete the attached EBS volume upon termination.

- **Stop Instance**

  When you stop an EC2 instance, the instance will be shut down and the virtual machine that was provisioned for you will be permanently taken away and you will no longer be charged for instance usage. The key difference between stopping and terminating an instance is that the attached bootable EBS volume will not be deleted. The data on your EBS volume will remain after stopping while all information on the local (ephemeral) hard drive will be lost as usual. The volume will continue to persist in its availability zone. Standard charges for EBS volumes will apply. Therefore, you should only stop an instance if you plan to start it again within a reasonable timeframe. Otherwise, you might want to terminate an instance instead of stopping it for cost saving purposes.

38. **Advantages of s3 service.**

- Reliable Security:
- All-time Availability:
- Very Low cost:
- Ease of Migration:
- The Simplicity of Management:

39. **Difference between TCP and UDP**

Basis	Transmission control protocol (TCP)	User datagram protocol (UDP)
Type of Service	TCP is a connection-oriented protocol. Connection-orientation means that the communicating devices should	UDP is the Datagram-oriented protocol. This is because there is no overhead for opening a connection,

	establish a connection before transmitting data and should close the connection after transmitting the data.	maintaining a connection, and terminating a connection. UDP is efficient for broadcast and multicast types of network transmission.
Reliability	TCP is reliable as it guarantees the delivery of data to the destination router.	The delivery of data to the destination cannot be guaranteed in UDP.
Error checking mechanism	TCP provides extensive error-checking mechanisms. It is because it provides flow control and acknowledgment of data.	UDP has only the basic error checking mechanism using checksums.
Acknowledgment	An acknowledgment segment is present.	No acknowledgment segments.
Sequence	Sequencing of data is a feature of Transmission Control Protocol (TCP). this means that packets arrive in order at the receiver.	There is no sequencing of data in UDP. If the order is required, it has to be managed by the application layer.
Speed	TCP is comparatively slower than UDP.	UDP is faster, simpler, and more efficient than TCP.
Retransmission	Retransmission of lost packets is possible in TCP, but not in UDP.	There is no retransmission of lost packets in the User Datagram Protocol (UDP).
Header Length	TCP has a (20-60) bytes variable length header.	UDP has an 8 bytes fixed-length header.
Weight	TCP is heavy-weight.	UDP is lightweight.
Handshaking Techniques	Uses handshakes such as SYN, ACK, SYN-ACK	It’s a connectionless protocol i.e. No handshake
Broadcasting	TCP doesn’t support Broadcasting.	UDP supports Broadcasting.
Protocols	TCP is used by HTTP, HTTPs, FTP, SMTP and Telnet.	UDP is used by DNS, DHCP, TFTP, SNMP, RIP, and VoIP.
Stream Type	The TCP connection is a byte stream.	UDP connection is message stream.
Overhead	Low but higher than UDP.	Very low.

40. Difference between region and availability zone.

AWS Regions are the actual geographic places in which the AWS data centers are located. An AWS Region has at least one Availability Zone and this number can go up to 6 in some of them.

Availability Zones are the actual AWS data centers that are located in these regions. AZs can have either a single or culmination of closely situated data centers.

41. Purchasing options of EC2 instances.

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

- **On-Demand Instances –** Pay, by the second, for the instances that you launch.
- **Savings Plans –** Reduce your Amazon EC2 costs by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.
- **Reserved Instances –** Reduce your Amazon EC2 costs by making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years.
- **Spot Instances –** Request unused EC2 instances, which can reduce your Amazon EC2 costs significantly.
- **Dedicated Hosts –** Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.
- **Dedicated Instances –** Pay, by the hour, for instances that run on single-tenant hardware.
- **Capacity Reservations –** Reserve capacity for your EC2 instances in a specific Availability Zone for any duration.

42. What is Website?

Website is the collection of web pages, different multimedia content such as text, images, and videos which can be accessed by the URL which you can see in the address bar of the browser. For example: http://www,amazon.com

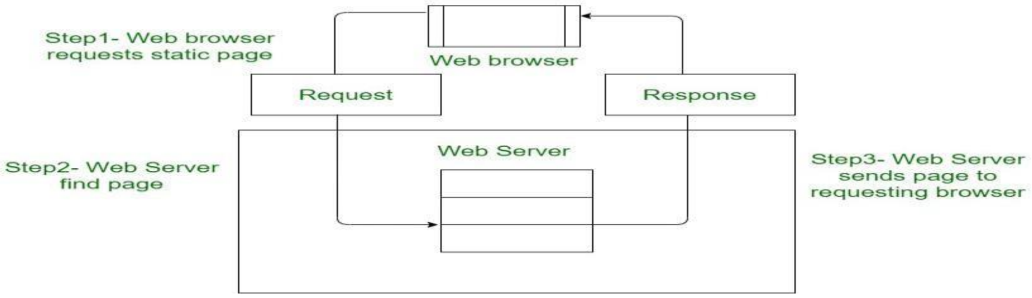
43. How to access Websites?

When we type a certain URL in a browser search bar, the browser requests the page from the Web server and the Web server returns the required web page and its content to the browser. Now, it differs how the server returns the information required in the case of static and dynamic websites.

Types of Websites

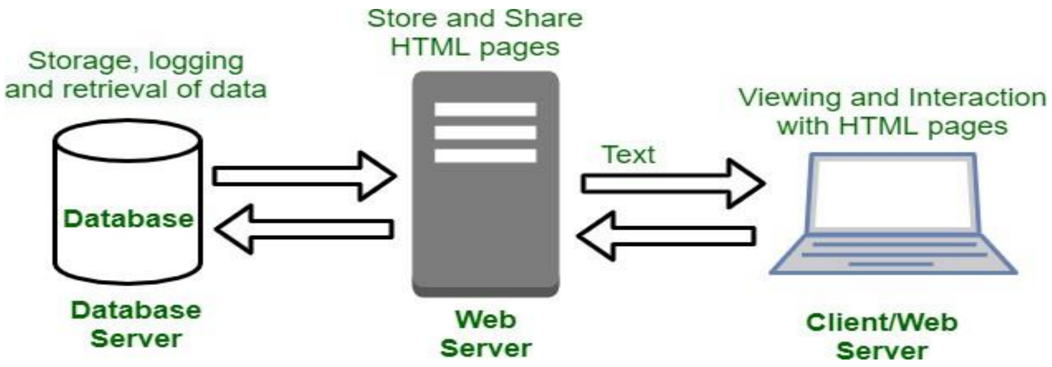
- Static Website
- Dynamic Website

I. **Static Website:** In Static Websites, Web pages are returned by the server which are prebuilt source code files built using simple languages such as HTML, CSS, or JavaScript. There is no processing of content on the server (according to the user) in Static Websites. Web pages are returned by the server with no change therefore, static Websites are fast. There is no interaction with databases. Also, they are less costly as the host does not need to support server-side processing with different languages.


Step1- Web browser requests static page
Web browser
Request
Response
Step2- Web Server find page
Web Server
Step3- Web Server sends page to requesting browser

Note: Static does not mean that it will not respond to user actions, These Websites are called static because these cannot be manipulated on the server or interact with databases (which is the case in Dynamic Websites).

II. Dynamic Website: In Dynamic Websites, Web pages are returned by the server which are processed during runtime means they are not prebuilt web pages but they are built during runtime according to the user's demand with the help of server-side scripting languages such as PHP, Node.js, ASP.NET and many more supported by the server. So, they are slower than static websites but updates and interaction with databases are possible.
Dynamic Websites are used over Static Websites as updates can be done very easily as compared to static websites (Where altering in every page is required) but in Dynamic Websites, it is possible to do a common change once and it will reflect in all the web pages.


Store and Share HTML pages
Storage, logging and retrieval of data
Viewing and Interaction with HTML pages
Text
Database
Database Server
Web Server
Client/Web Server

44. What is AMI?

An Amazon Machine Image (AMI) is a template that contains a software configuration, launch permission and a block device mapping that specifies the volume to attach to the instance whence launched. (for example, an operating system, an application server, and applications). From an AMI, you launch an instance, which is a copy of the AMI running as a virtual server in the cloud.

Types of AMI

As an AWS user, you can choose your AMI on the basis of the following parameters:

a) **Operating System**

You can choose an AMI on the basis of the supported operating system (or OS) like Windows or Linux.

b) **32-bit or 64-bit Architecture**

This parameter is based on the architecture of your selected OS.

c) **Region**

This parameter is based on the selected region of the Amazon machine image that comprises regions, availability zones, and local zones. Each region operates in different geographical regions and is independent of each other.

d) **Storage (EBS or Instance store)**

This AMI parameter is based on the storage of the root device. Based on storage, AMIs are categorized as either of the following two types namely:

- **EBS-backed instances:** In this case, the root device for an AWS instance – launched using AMI – is an Amazon EBS volume that has been created from Amazon EBS.
- **Instance store-backed instances:** In this case, the root device for an AWS instance – launched using Ami – is an Amazon instance store volume that has been created from an Amazon S3 template.

45. Difference between EBS and instance store

Instance Store:

- ➢ EC2 instances using instance store as root device volumes are called as Instance Store backed root device volumes. Instance store volume can be created from a template stored in the S3 bucket.
- ➢ An instance store provides temporary block-level storage for your instance. This is the disk that is physically attached to your host computer.
- ➢ Instance Store are sometimes also referred to as Ephemeral storage, that are suitable for buffer, cache, scratch data & other temporary contents.
- ➢ Data on an instance store volume persists only during the life of the associated instance; if an instance is stopped or terminated, any data on instance store volumes is lost.
- ➢ Since the instance stores aren’t persistent across reboots, they are sometimes also referred to as Ephemeral storage.
- ➢ Instance storage are the closest available memory to the instance, hence offer the lowest latency.

Instance store volumes are complicated and its best to gauge their limitations before planning to use them. Few limitations listed below are:

- In order to deliver very high random, I/O performance, instance store volumes only run-on solid-state drives (SSD).
- Instance store backed Instances cannot be upgraded
- Taking a snapshot or AMI of an instance store volume is not as straightforward as taking snapshot of EBS volume.
- Data in the instance store can be lost if the underlying disk drive fails or if the instance stops, or if the instance terminates.

EBS Backed Root Volumes:

- An “EBS-backed” EC2 instance means that the root device for an EC2 instance launched from the AMI is an EBS volume created from an AWS EBS snapshot.
- AWS EBS is high performance block storage designed for use for EC2 instances handling throughput and transaction intensive workloads. These workloads may include relational, non-relational databases, containerized applications, file systems etc.
- EBS storage is storage on a remote network connected SAN or NAS (Network Attached Storage)
- Volume persists independently from the running life of an instance. After an EBS volume is attached to an instance, you can use it like any other physical hard drive.
- EBS volume can be detached from one instance and attached to another instance, supports encryption, and is also replicated across multiple availability zones to provide high availability & durability.
- Data on EBS stores persists over the reboots of the EC2 instances.
- EBS backed instances can help you save money as they can be turned off and when not being used

46. What is instance?

An instance is a virtual server in the cloud. Its configuration at launch is a copy of the AMI that you specified when you launched the instance.

47. What is template?

A launch template is an Amazon Elastic Compute Cloud (EC2) feature that reduces the number of steps that are required to create an AWS instance by capturing all launch parameters within one resource.

48. What Is an EBS Snapshot?

An EBS snapshot is a point-in-time backup of your EBS volume. It is a “copy” of the data on your EBS volume.

49. Is an EBS Snapshot a Full or Incremental Backup?

An EBS snapshot is actually both a full backup and an incremental backup.

When an EBS snapshot is created, only the data on the EBS volume that has changed since the last EBS snapshot is stored in the new EBS snapshot. In this way, it’s an incremental backup.

Internally, the EBS snapshots chain together.

When an EBS snapshot is used to restore data, all data from that EBS snapshot can be restored as well as the data from the previous snapshots. In this way, the snapshot is a full backup.

50. Is It Safe to Delete Old EBS Snapshots?

Yes. You can safely delete old EBS snapshots. New EBS snapshots will continue to restore properly.

When you delete an old EBS snapshot, behind the scenes, AWS will consolidate the snapshot data. It will move valid data forward to the next EBS snapshot and it will discard invalid data.

51. Difference between ARM and X86

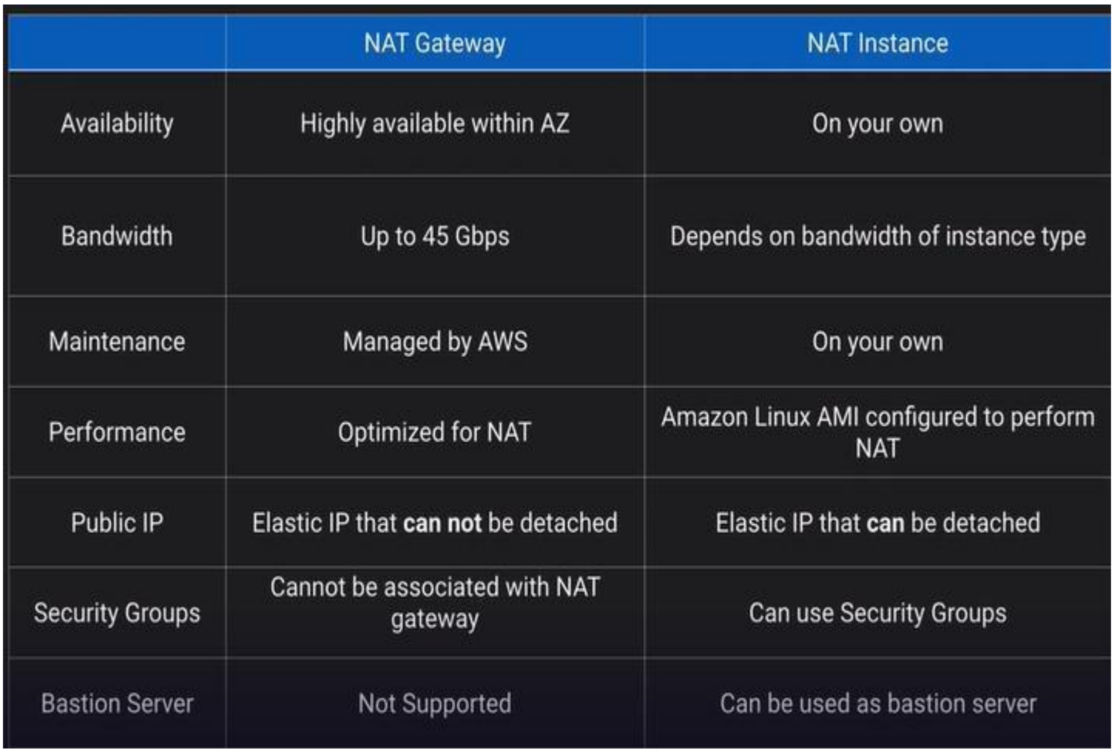
ARM	X86
Uses Reduced Instruction Set Computing Architecture (RISC).	Uses Complex Instruction Set Computing Architecture (CISC).
Executes single instruction per cycle.	Executes complex instruction at a time, and it takes more than a cycle.
Optimization of performance with Software focused approach.	Hardware approach to optimize performance.
Requires less registers, more memory.	It uses more registers and less memory
Pipelining of instructions is a unique feature.	Less pipelined.
Faster Execution of Instructions reduces time.	Time to execute is more.
Complex addressing is managed by software.	Inherently designed to handle complex addresses.
Compiler plays a key role in managing operations.	The micro program does the trick.
Multiple Instructions are generated from a complex one and executed individually.	Its Architecture is capable of managing complex statement execution at a time.
Managing code expansion is difficult.	Code expansion is managed easily.
Decoding of instruction is handled easily.	Decoding is handled in a complex way.
Uses available memory for calculations.	Needs supplement memory for calculations.
Deployed in mobile devices where size, power consumption speed matters.	Deployed in Servers, Desktops, Laptops where high performance and stability matters.

52. AWS NAT Instances & NAT Gateways

A NAT (Network Address Translation) instance

It's, like a bastion host, an EC2 instance that lives in your public subnet. A NAT instance, however, allows your private instances outgoing connectivity to the internet while at the same time blocking inbound traffic from the internet. Many people configure their NAT instances to allow private instances to access the internet for important operating system updates. As I've discussed previously, patching your OS is an important part of maintaining instance level security.

NAT Gateways provide the same functionality as a NAT instance; however, a NAT Gateway is an AWS managed NAT service. As a result, these NAT Gateways offer greater availability and bandwidth and require less configuration and administration.



	NAT Gateway	NAT Instance
Availability	Highly available within AZ	On your own
Bandwidth	Up to 45 Gbps	Depends on bandwidth of instance type
Maintenance	Managed by AWS	On your own
Performance	Optimized for NAT	Amazon Linux AMI configured to perform NAT
Public IP	Elastic IP that **can not** be detached	Elastic IP that **can** be detached
Security Groups	Cannot be associated with NAT gateway	Can use Security Groups
Bastion Server	Not Supported	Can be used as bastion server

53. What are AWS CloudWatch?

CloudWatch collects monitoring and operational data in the form of logs, metrics and events, and visualizes it using automated dashboards so you can get a unified view of your AWS resources, applications and services that run in AWS and on-premises.
Using AWS CloudWatch, you can monitor your AWS account and resources and generate a stream of events or trigger alarms and actions for specific conditions.

54. AWS Multi-Factor Authentication (MFA)

As you know storing sensitive information in the cloud is vulnerable to hackers and viruses, and to overcome this your account needs to be secured. For increased security, AWS recommends that you configure AWS Multi-Factor Authentication (MFA) to help protect your AWS resources.
In brief, MFA = Password you know + Security Device you own

AWS Multi-Factor Authentication (MFA) is a simple best practice that adds an extra layer of protection on top of your users name and password. With MFA enabled, when a user signs in to an AWS Management Console, they will be prompted for their user name and password (the first factor what they know), as well as for an authentication code from their AWS MFA device (the second factor what they have). Taken together, these multiple factors provide increased security for your AWS account settings and resources.

55. Why AWS MFA is Required

- Users have access to your account and can possibly change configurations and delete resources in your AWS account, so to overcome this it is required
- If you want to protect your root accounts and IAM user.

- Even if the password is stolen or hacked, the account is not compromised.
- When you enable this authentication for the root user, it affects only the root user credentials. IAM users in the account are distinct identities with their own credentials, and each identity has its own MFA configuration.

56. MFA Device Options In AWS

The following are the MFA device options in AWS:

- **Virtual MFA devices**
  - A software app that runs on a phone or other device and emulates a physical device.
  - The device generates a six-digit numeric code based upon a time-synchronized one-time password algorithm.
  - Virtual MFA might not provide the same level of security as FIDO2 devices or hardware MFA devices.
- **FIDO security key**
  - A device that you plug into a USB port on your computer.
  - FIDO2 is an open authentication standard hosted by the FIDO Alliance.
  - When you enable a FIDO2 security key, you sign in by entering your credentials and then tapping the device instead of manually entering a code.
- **Hardware MFA device**
  - A hardware device that generates a six-digit numeric code based upon a time-synchronized one-time password algorithm.

Note: - *If your AWS account root user multi-factor authentication (MFA) device is lost, damaged, or not working, you can recover access to your account. IAM users must contact an administrator to deactivate the device.*

57. Assume Role

Assuming a role means obtaining a set of temporary credentials which are associated with the role and not with the entity that assumed the role.

58. RDS vs EC2 differences

Amazon RDS enables you to run a fully featured relational database while offloading database administration. Whereas, for more control and flexibility, EC2 will be better for your relational database.

If you want an automated and cost-effective solution, go for RDS. With Amazon RDS, AWS will take care of your database from end-to-end. AWS offers an automated process for configuring, managing, maintaining, and securing. Whereas, for more control and flexibility, EC2 will be better for your relational database.

	Amazon RDS	**Amazon EC2 Relational Databases**
Administration	AWS takes full responsibility for your database. The entire process of configuration, management,	EC2 gives you full control over your database, OS and software stack. It allows

	maintenance, and security is automated by AWS.	you to hire your own database administrators.
Availability	RDS is a highly available relational database. It automatically creates a primary DB instance and replicates the data side by side to a standby instance in a different Availability Zone.	With EC2, you have to set up your database for high availability.
Scalability	Scaling up your database is comparatively easier with Amazon RDS. This can be done by adding replicas. It allows you to easily configure read replicas or set up synchronous replication across availability zones for enhanced performance, availability, and durability.	With EC2, you have to setup such architecture (Availability Groups, Sharding, and more) manually with help of other EC2 instances and load balancer.
Backups	RDS offers automated backups. Plus, you can get snapshots on-demand and keep them with you as long as you wish to.	With EC2, you have to take care of backup.
Performance	RDS offers Provisioned IOPS or PIOPS to achieve fast, predictable, and consistent Input/Output performance.	EC2 allows you to meet unique performance, replication, archival or DR requirements by giving you the required flexibility. You can choose the EBS (SSD) volume as per your need
Storage	In RDS, you get 3 types of storage options: • General Purpose SSD: It offers cost-effective storage that is ideal for a broad range of workloads. These volumes deliver single-digit millisecond latencies and the ability to burst to 3,000 IOPS for extended periods of time. • Provisioned IOPS: It is designed to meet the needs of I/O-intensive workloads, particularly database workloads, that require low I/O latency and consistent I/O throughput. • Magnetic: It supports magnetic storage for backward compatibility.	In EC2, you get the following: • You can get up to 8000 IOPS and 800 MBPS with provisioned IOPS and the right EC2 instance (It depends on instance type). • You can use EBS RAID and striping configurations for higher and better performance.
Compatibility	RDS supports Aurora, SQL Server, MySQL, MariaDB, PostgreSQL, and Oracle.	With EC2, you can configure any database you want.
Control	With RDS, you don’t have control over the system.	EC2 offers complete control over the system. Complete control is one of the key benefits of EC2.

Security	RDS offers encryption at rest and in transit. Data that is encrypted at rest includes the underlying storage for DB instances, Read Replicas, its automated backups and snapshots.	In EC2, EBS volumes are encrypted to protect your data, both at rest and in motion. This is majorly beneficial when traveling from EBS volume to EC2 instance.

If you want an automated and cost-effective solution, go for RDS. Whereas, for more control and flexibility, go for EC2.

RDS has a little edge over EC2:
1. Optimized DB Solutions & Replication: There is no need to manually set up database mirroring and failover clusters since you get highly optimized database solutions and synchronous multi-AZ replication.
2. Outsource tasks: You can outsource tasks like database provisioning, security, and updating versions. No need of DBAs.
3. Focus on important tasks: It allows you to focus on tasks like schema optimization and performance tuning.
4. Automatic Backups: In case of a disaster, RDS manages your backups automatically.

However, EC2 is also preferred by few people:
1. Full control: It gives you maximum control over software stack, database, and OS.
2. Database Admins: Manage your database by looking after clustering, replication, and backups.
3. More features: You can use SQL Server features that aren’t currently supported by Amazon RDS. (Now RDS supports SSRS)
4. High Performance: It allows you to exceed your maximum database size and performance needs.

59. How to check whether my user data passing to EC2 instance is executed or not?

You can verify using the following steps:

- SSH on launch EC2 instance.
- Check the log of your user data script in:
- /var/log/cloud-init.log and
- /var/log/cloud-init-output.log

You can see all logs of your user data script, and it will also create the **/etc/cloud** folder.

60. What is Target Group?

The target group lets to know the load balancer, where to direct the traffic to EC2 instances, fixed IP addresses or Lambda functions, out of other resources. While we create a load balancer, we create single or multiple listeners and set the listener rules to direct the traffic to a single group.

61. What is stickiness?

Stickiness is a term that is used to describe the functionality of a load balancer to repeatedly route traffic from a client to a single destination, instead of balancing the traffic across multiple destinations.

62. What is t2micro instance?

T2 instances are Burstable Performance Instances that provide a baseline level of CPU performance with the ability to burst above the baseline.

63. Amazon EC2 instance store

An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer. Instance store is ideal for temporary storage of information that changes frequently, such as buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers.

64. What's the difference between cloud and virtualization?

It’s easy to confuse virtualization and cloud, particularly because they both revolve around creating useful environments from abstract resources. However, virtualization is a technology that allows you to create multiple simulated environments or dedicated resources from a single, physical hardware system, and clouds are IT environments that abstract, pool, and share scalable resources across a network. To put it simply, virtualization is a technology, where cloud is an environment.

	Virtualization	**Cloud**
Definition	Technology	Methodology
Purpose	Create multiple simulated environments from 1 physical hardware system	Pool and automate virtual resources for on-demand use
Use	Deliver packaged resources to specific users for a specific purpose	Deliver variable resources to groups of users for a variety of purposes
	Virtualization	**Cloud**
Configuration	Image-based	Template-based
Lifespan	Years (long-term)	Hours to months (short-term)
Cost	High capital expenditures (CAPEX), low operating expenses (OPEX)	Private cloud: High CAPEX, low OPEX Public cloud: Low CAPEX, high OPEX
Scalability	Scale up	Scale out
Workload	Stateful	Stateless
Tenancy	Single tenant	Multiple tenants

65. What is subnet (subnetwork)?

A subnet, or subnetwork, is a segmented piece of a larger network. More specifically, subnets are a logical partition of an IP network into multiple, smaller network segments.

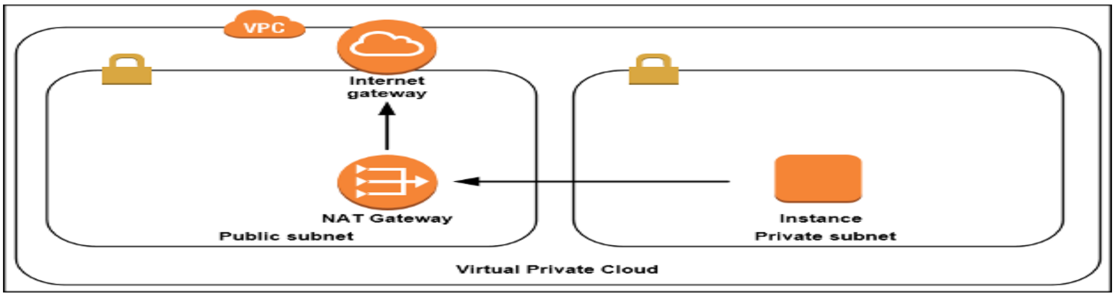
66. What is a Public subnet?

A public subnet is a subnet that is associated with a route table that has a route to an Internet gateway. This gateway connects the subnet(work) to the Internet and to other AWS services. Thus, the resources in that subnet able to communicate with the Internet.

67. What is a Private subnet?

A private subnet is a subnet that is associated with a route table that doesn't have a route to an internet gateway. Resources in a private subnet cannot communicate directly with the Internet, and vice versa.

68. What is the difference between public and private subnets?



The instances in the public subnet can send outbound traffic directly to the Internet with the help of Internet Gateway, whereas the instances in the private subnet can't because we are not attaching Internet Gateway to the Private Subnets.
Instead, the instances in the private subnet can access the Internet by using a Network Address Translation (NAT) gateway that resides in the public subnet.

Instances or resources living in the Private Subnet will be safer than Public Subnet because any traffic initiated from the internet cannot reach directly to the endpoints in Private Subnet, but can reach in Public Subnet.

69. Types of users in AWS

- Root user
- IAM user
- Federated user

70. What are trusted policies in AWS?

This policy can define which principals can assume the role, and under which conditions. This is sometimes referred to as resource-based policy

71. Use of primary and secondary IP address.

Alias/Secondary IP Address
In the context of IP Addressing, Aliasing refers to the process of creating and configuring multiple IP addresses on a single Network Interface. By having aliased IP addresses, you can create multiple connections to a network for an individual node on a network where each connection can serve for a different purpose.

In GCP we can configure a primary and at the same time, a secondary CIDR Range as a part of a subnet:

- Configure 10.1.0.0/16 as a primary CIDR Range
- Configure 10.2.0.0/20 as a secondary CIDR Range

Here, primary IP Address of Virtual Machine is allocated from primary CIDR range, whereas an alias IP range, 10.2.1.0/24, is allocated in VM from secondary CIDR range. Some most common applications of Secondary IP Address include:

- Secondary IP address is a good option to be used in case of hosting sites on a dedicated IP with SSL certificate because if you are using HTTPS, you need an IP per website.
- If you wish to split resources across your IP's so that you can apply different firewalling for access control or put in some QoS rule, then various services can be assigned to different IP's.
- Using a Secondary IP Address can be useful for controlling traffic easily depending on how you segregate DNS names/services between them.

NOTE: - Secondary addresses do not support DHCP.

72. SSL Certificate

An SSL certificate is a digital certificate that authenticates a website's identity and enables an encrypted connection. SSL stands for **Secure Sockets Layer**, a security protocol that creates an encrypted link between a web server and a web browser.

Companies and organizations need to add SSL certificates to their websites to secure online transactions and keep customer information private and secure.

In short: SSL keeps internet connections secure and prevents criminals from reading or modifying information transferred between two systems. When you see a padlock icon next to the URL in the address bar, that means SSL protects the website you are visiting.

Since its inception about 25 years ago, there have been several versions of SSL protocol, all of which at some point ran into security troubles. A revamped and renamed version followed — TLS (Transport Layer Security), which is still in use today. However, the initials SSL stuck, so the new version of the protocol is still usually called by the old name.

73. How do SSL certificates work?

SSL works by ensuring that any data transferred between users and websites, or between two systems, remains impossible to read. It uses encryption algorithms to scramble data in transit, which prevents hackers from reading it as it is sent over the connection. This data includes potentially sensitive information such as names, addresses, credit card numbers, or other financial details.

74. Amazon Route53

Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. It is basically designed for developers and corporate to route the end users to Internet applications by translating human-readable names like www.geeksforgeeks.org into the numeric IP addresses like 192.0.1.1 those computers use to connect to each other. You cannot use Amazon Route 53 to connect your on-premises network with AWS Cloud.

75. Functions of Route53

- If a web application requires a domain name, Route53 service helps to register the name for the website (i.e domain name).
- Whenever a user enters the domain name, Route53 helps to connect the user to the website.
- If any failure is detected at any level, it automatically routes the user to a healthy resource.
- Amazon Route 53 is cost effective, secure and scalable.
- Amazon Route 53 is flexible, highly available and reliable.

76. Methodologies related to Route53

- **Records:** Records are created to route internet traffic to the resources. They are the objects present in the hosted zone which determines how the internet traffic has to be routed for a domain name so that it finally reaches the resources. The name of each record in a hosted zone must end with the name of the hosted zone.

  Types of records

  a) **SOA (Start of Authority Records)**

  Basic SOA stores information about below things.

  - Name of Server that supplied the data for zone.
  - The administrator of that zone & current version of data file.

  b) **NS Record (Name Server Records)**

  NS records is basically your name server records which are used by top level domain servers to direct traffic to content DNS server which contains the authoritative records.

  NOTE - *So whenever we create a hosted zone in Route53, Two types of records automatically created, one is SOA & second is NS.*

  Before we head over to important type of records used in AWS, let's talk about one important concept TTL(Time to Live)

  TTL (Time to Live):

  TTL is mandatory for each DNS record. So TTL is length that a DNS record is cached on either the resolving server or user own Laptop. The Lower the TTL, the faster changes to DNS records. Whenever you created record set, you need to define TTL for it.

c) **A Record (URL to IPv4)**
The “A” record stands for Address record. The A record is used by computer to translate the name of the domain to an IP address.
E.g.: (http://medium.com might point to http://126.78.98.90)

d) **CNAME (Canonical Records- URL to URL)**
CNAME Points a URL to any other URL. (gaurav.gupta.com => gkg.example.com), We use it only for Non-Root Domain (aka. something.mydomain.com)

e) **Alias Record:**
Alias record points a URL to an AWS Resource, Alias record are used to map resource record sets in your hosted zone to Elastic Load Balancer, CloudFront or S3 Buckets websites.

f) **AAAA: (URL to IPv6)**
An AAAA record maps a domain name to the IP address (Version 6) of the computer hosting the domain. An AAAA record is used to find the IP address of a computer connected to the internet from a name.

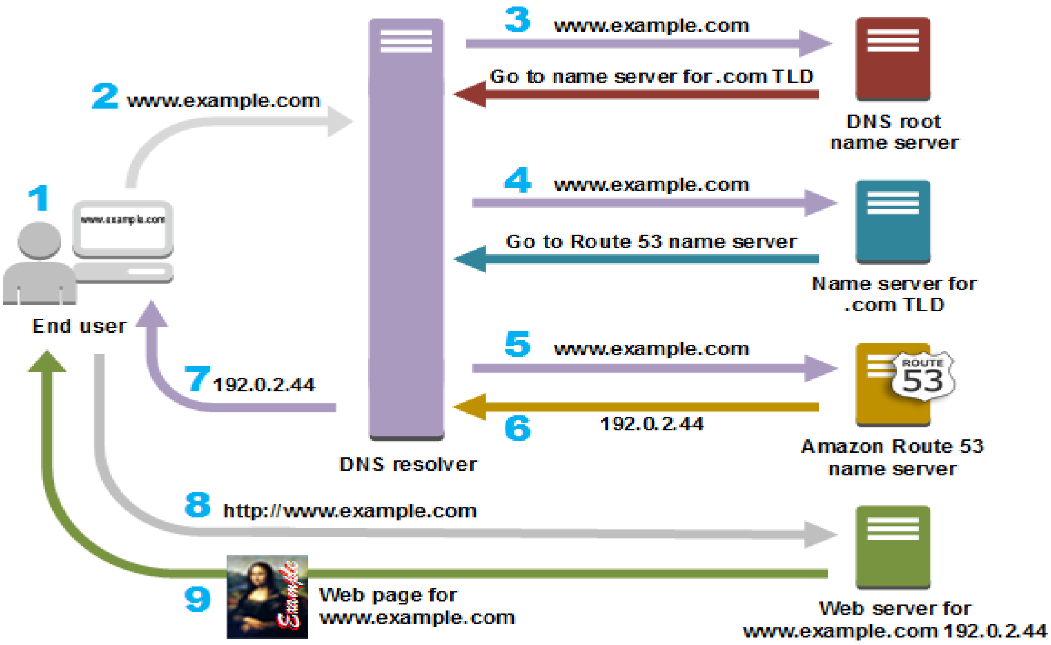
g) **MX Record (Main Exchange Record)**
A mail Exchanger record (MX record) specifies the mail server responsible for accepting email messages on behalf of a domain name. It is a resource record in the Domain Name System (DNS). It is possible to configure several MX records, typically pointing to an array of mail servers for load balancing and redundancy.

- **Hosted zone:** When the domain name is registered, Route53 creates a public hosted zone that has the same name as the domain name. It is a collection of records that contains information about how to route traffic of its domains and all of its subdomains.
    - Create a hosted zone with either a public hosted zone or a private hosted zone:
        - Public Hosted Zone – for routing internet traffic to the resources for a specific domain and its subdomains
        - Private hosted zone – for routing traffic within a VPC
    - Create records in the hosted zone
      Records define where to route traffic for each domain name or subdomain name.
      Name of each record in a hosted zone must end with the name of the hosted zone.
- **DNS query:** It is a request for information sent from DNS client to the DNS server.
- **Alias records:** Alias records helps in routing internet traffic to AWS resources like S3 bucket, Amazon CloudFront, etc. It is created at the top node of the DNS namespace.
- **Name servers:** They are the servers in the DNS that translates the domain name into IP address so that internet traffic can be routed to the resources.
- **DNS failover: A** method for routing the traffic from unhealthy resources to healthy resources, whenever a failure is detected.

- **Routing policy:** Routing policy determines how Amazon Route53 responds to queries.

77. How Does DNS Route Traffic To Your Web Application?

The following diagram gives an overview of how recursive and authoritative DNS services work together to route an end user to your website or application.



78. Types of Routing Policy

- **Simple routing policy**: It is a simple Route53 routing technique that can be used to route internet traffic to a single resource. For example; Web server to a website. Using this, routing multiple records with the same name cannot be created but multiple values (such as multiple IP addresses) can be specified in the same record.
- **Failover routing policy:** Whenever a resource goes unhealthy, this policy allows to route the traffic from unhealthy resource to healthy resource.
- **Geolocation routing policy:** This routing policy routes the traffic to resources on the basis of the geographic location of the user. Geographic locations can be specified by continent, country, or state. For example; A person residing in France will be redirected to the website in the French language while a person from the US will be redirected to the website in the English language.
- **Geoproximity routing policy:** It routes traffic on the basis of the geographical location of the user and the type of content user wants to access. The user can optionally shift traffic from resources at one location to resource at another location. Using this policy, a user can shift more traffic to one location compared to another location by specifying a value known as bias.
- **Latency routing policy:** If a website has to be hosted in multiple regions, then a latency-based routing policy is used. To improve performance for the users, this policy helps in serving requests from the AWS region that provides the lowest latency. To use this policy the latency records for the resources are created in multiple AWS regions.

- **Multivalue routing policy:** It is used when users want Route53 to return multiple values in response to DNS queries. It first checks the health of resources and then returns the multiple values only for the health resources.
- **Weighted routing policy:** This routing policy routes traffic to multiple resources with a single domain name according to the proportion decided by the user.

79. What type of encryption does S3 use?

Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256) GCM, to encrypt your data.

80. What is the difference between source ip and destination ip?

- **Source IP** is the IP (Internet Protocol) address of the device sending the IP packet (the IP unit of data transfer).
- **Destination IP** is the IP address of the device to which the packet is being sent.

It means the following,

Local Network: If you are sharing a file from the server to your PC in your office, then your PC's IP is the source IP and the server's IP is destination IP.

Internet: If you are a streaming a video from a music website from your PC in your office then your internet routers WAN IP is the source IP and music website's IP is destination IP.

81. Why we create inline policy in AWS?

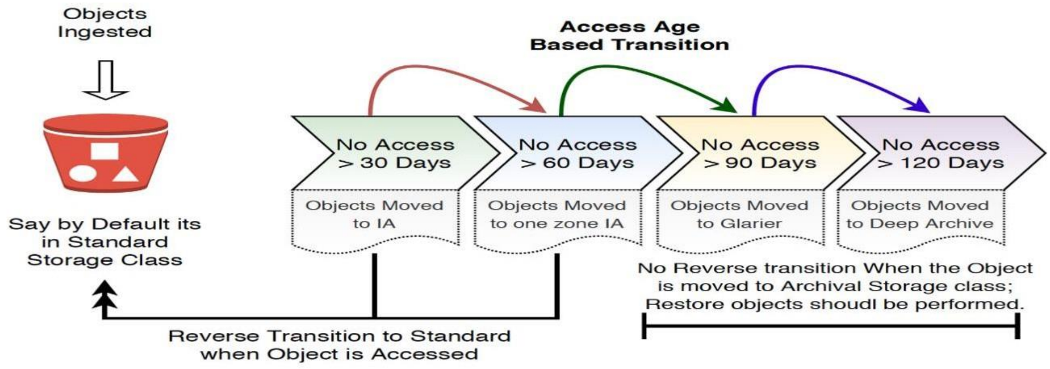
Inline policies are useful if you want to maintain a strict one-to-one relationship between a policy and the identity to which it is applied.

82. Amazon S3 life cycle management

Cost of data (Objects) storage is important to consider when making decisions to use right storage technology. Amazon S3 Lifecycle allows to configure a lifecycle for stored objects on S3, to optimize the cost. A Lifecycle configuration is a set of rules that define actions applied to a group of objects.

Types of Lifecycle configuration Actions:

- **Transition actions:** Moving objects from one storage class to another storage class. Each storage class has a different cost associated with it.
- **Expiration actions:** When objects expire after a span of time (say 30 days,60 days, etc). Amazon S3 deletes expired objects on your behalf.


Objects Ingested
Access Age Based Transition
No Access > 30 Days
No Access > 60 Days
No Access > 90 Days
No Access > 120 Days
Objects Moved to IA
Objects Moved to one zone IA
Objects Moved to Glarier
Objects Moved to Deep Archive
Say by Default its in Standard Storage Class
No Reverse transition When the Object is moved to Archival Storage class; Restore objects shoudl be performed.
Reverse Transition to Standard when Object is Accessed

We also support reverse transition for Non-Archival Storage class (Glacier or Deep Archive), by moving the object from either “STANDARD_IA” or “ONEZONE_IA” to “STANDARD” storage class when the object is accessed. This reverse transition is performed instantly once object is accessed.

83. Can we attach single EBS to multiple EC2 instances?

With the new AWS EBS Multi-Attach option, users can now attach a single EBS volume with a maximum of 16 **Nitro based Amazon EC2 instances**.

Amazon EBS Multi-Attach enables you to attach a single Provisioned IOPS SSD (io1 or io2) volume to multiple instances that are in the same Availability Zone.

84. How many volumes can you attach to a single EC2 instance?

For example, if you have no additional network interface attachments on an EBS-only instance, **you can attach up to 27 EBS volumes** to it. If you have one additional network interface on an instance with 2 NVMe instance store volumes, you can attach 24 EBS volumes to it.

85. What is S3 FUSE?

s3fs is a FUSE filesystem that allows you to mount an Amazon S3 bucket as a local filesystem. It stores files natively and transparently in S3 (i.e., you can use other programs to access the same files).

86. How many elastic IPS can be connected to an instance?

five Elastic IP addresses

87. Amazon Resource Names (ARNs)

Amazon Resource Names (ARNs) uniquely identify AWS resources. We require an ARN when you need to specify a resource unambiguously across all of AWS, such as in IAM policies, Amazon Relational Database Service (Amazon RDS) tags, and API calls.

ARN format

The following are the general formats for ARNs. The specific formats depend on the resource. To use an ARN, replace the italicized text with the resource-specific information. Be aware that the ARNs for some resources omit the Region, the account ID, or both the Region and the account ID.

arn:partition:service:region:account-id:resource-id
arn:partition:service:region:account-id:resource-type/resource-id
arn:partition:service:region:account-id:resource-type:resource-id

partition

The partition in which the resource is located. A partition is a group of AWS Regions. Each AWS account is scoped to one partition.

- The following are the supported partitions:
- aws - AWS Regions
- aws-cn - China Regions
- aws-us-gov - AWS GovCloud (US) Regions