AWS

- The applications in the AWS services are created in a private network using the Amazon virtual private cloud(VPC). Any resource you put in the VPC will be public.
- The backend code of the application is done on the amazon Elastic compute cloud(EC2) and the EC2 also offers virtual machines.
- The data will be stored in a virtual database of AWS called the Relational Database service.
- The amazon simple storage service S3 is used to store unlimited amount or large files of data.
- The AWS have a practice know as “Availability zone” in which an other data center is connected to the main data center, if there is any interruption in the service of the main data center then the secondary data center in the availability zone will be used in action.
- And the cluster(group of) AZ’s is known as the Regions.

The four factors of region consideration :

1. Compliance - If you have any restrictons that your data must consider in this location, or if you have any data boundaries that my data should not cross this region then you must consider the regions that reside in your country or state.
2. Latency - More the distance of the data center from the user the more latency.
3. Pricing - Prices of the regions
4. Service availability - Availability of service

- To reduce the latency using the cache locations or the edge locations is necessary, this reduce the loading time of the site.

> 💡 **API - application programming interface.**

Ways to interact with AWS API

1. AWS management console - The GUI in which a user can perform actions by logging in into the AWS account via a web browser.
2. AWS Command line interface - A terminal or a command line interface where a user uses the aws service by using a Syntax that is predefined.
   We can automate the tasks using the AWS CLI
3. AWS software development kits - In helps in linking the codes that we write in various programming languages, and to access the AWS services with the code.

Security in AWS

AWS follows the shared responsible model.

- This makes the user also responsible for the Security of the cloud.
- The AWS provides security of the cloud, that is the hardware of the cloud, AZ's, Region, Networking, databases and other aspects that are important for running the cloud.
- The User are responsible for managing the access, that is who can access there data and cloud, OS and firewall configuration and etc.
- The AWS root user must provide the username and the password to login into the AWS management console and use the services
- AWS provides you second set of authentication, it is using the access keys, it is uses an access key and a secret access key, you must know the both access key and secret access key to login to your account.
- The AWS CLI is used in the case of login via access key.
- Using of the MFA(Multi Factor Authentication) is the best practice to make your account more secured.
- As using the single Factor authentication isn't that reliable the MFA uses 3 types of authentication layers that protects your aws root account from threat actors.
    - It requires Username and password

- OTP that comes directly to your registered device.
- Your fingerprint or face ID.

AWS Identity and access management

The IAM manages the login credentials and permission to the AWS account, also can manage the credentials used to sign API calls made to AWS services.

- IAM policies are also used to manage the permissions of the users, that is if they are allowed to do any task or not.
- IAM policies can also be attached to a group of users.
- The best way is, create your AWS root user account with an MFA authentication and give an IAM user the admin permissions, then log in with the IAM user account and manage the other IAM policies for other users and groups, and you cannot apply a policy to a root user.
- IAM also supports MFA.
- The IAM Groups can have any number of users, A user can belong to many groups, but a group cannot belong to a group.
- If you have many employees in your organization and if a few employees have more than one role than there is now need to create as many as IAM users for the employee, just use the Identity provider.
- The AWS IAM identity center is an idp that lets users to sign in using one login credentials, and then, and the remaining tasks of assigning the user policies and everything is same.

Polices

- The **IAM policies** are mostly written in the JSON form, these policies are those which grant permissions to the users where they are allowed or denied to access any AWS service
- In the action section of IAM policies, “*” this asterisk symbol is known as **wildcard,** it means full access to any actions.

Role based

- An IAM role is assigned for any temporary access to the AWS credentials.
- IAM roles does not have any Login credentials, they only login programmatically.
- Roles are used to send API calls to other AWS services, by using temporary Credentials.

Compute

There are 3 types of compute options:

1. Virtual machines - Elastic Compute cloud
2. Container services - Elastic container service, Elastic kuberrnetes service.
3. Serverless

Elastic cloud compute(EC2)

- The EC2 is an infamous AWS services that allows users to host virtual machines.
- An Instance is nothing but a single virtual machine.
- EC2 emulates a physical server and it creates virtual machines that use the HTTP server to run your application. And the virtual machines need the use of hypervisor to create an run virtual machines on your host machine.
- But behind the scenes AWS also manages the virtual machines and the hypervisor layer.
- An AWS user can select his required instance and can pay for the services he used at the end of the billing cycle.
- The instance type G is mostly used in order for 3D visualization and video enhancement.
- M5 provides a balance of resources.
- EC2 is scalable that is, when you are creating an instance it also asks you the size of the instance that you need, and the size can be changed after words.

- while creating an instance we select the AMI that is the properties that our virtual machine should have such as the OS, and storage mappings.
- The process includes, after creating an instance the AWS allocates virtual machine using the hypervisor and the, AMI is then allocated or copied to the Root volume, which contains the image that is used to boot the volume.
- We can create an another instance with same properties using the instance that we created before.
- Using the terminate protection we can save the data for a limited amount of time before it get deleted.
- EC2 instances are a combination of virtual processors, and in a few times AWS also provides GPU's

> 💡 C5.large - here 'C' refers to the instance family and the 5th generation, and large denotes the size of the instance.

- The EC2 instances live in the amazon VPC.
- **High availability -** Using 10 small instances is better than using 2 large instances.
- **EBS - elastic block store.**
- The EC2 internal storage is called internal storage and the external storage is called the Amazon Elastic block storage EBS.

Containers

- Containers are used in order to pack all the things such as packages and dependencies everything in a whole package.
- They are independent of platforms.
- They are faster as compared to the virtual machines.
- And to maintain and manage this containers AWS uses orchestration container services like, ECS and EKS.

- **ECS maintains containers across the many EC2 instances**, if there are more than one containers and more than one instance that are need to be managed then that is called a **Cluster.**
- To run and mange the containers you must install the **ECS container agent**.
- to prepare an application to run on the ECS you must define a text file in the JSON format, that describes one or more containers.
- The EKS is similar to ECS but the differences arise in the naming of few tasks.
- **AWS Fargate** is used to manage the container service in a serverless compute manner.
- We can store our container images in amazon ECR Elastic compute registry.

Serverless

This eases up the tasks that has to be done by the users.

- The serverless compute types are more convenient to the user as AWS take care of maintaining and managing of the servers and instances and other, the only the user must manage is the IAM.
- In the serverless compute model there is no need of the user to manage patches, and managing OS and

AWS Lambda

The AWS lambda is a popular serverless compute service.

- Lambda allows you to package and upload your code to the lambda and make an lambda function, and then the lambda function wont be running all the time.
- The lambda function is only run when a trigger is detected, the code is automatically run in the AWS, where lambda can process any code with a runtime of under 15 minutes.
- Lambda scales it self based on the triggers than are incoming, it can handle many triggers at single time by allocating each trigger an separate environment.
- You won’t be billed for the code that isn’t using.

Pricing in AWS

You pay what you use.

There a re three main purchasing options in AWS.

1. On-demand - this model is used when you don't want your servers to be working all the time, that is you only want the servers to be worked on-demand, and you will only billed for the time your instance are running, and the billing will be stopped if the instances are stopped.
2. reserved - This model is used if you need yo the AWS services through out 24/7 for a period of time say, 1,2,3 or many years.
3. spot instances - this model is used as the AWS allows you to pay for the spots or space that is free up in the cloud and set a maximum pay for it, and if the place in the could is needed for an on-demand customer AWS warns and interrupts you before 2 minutes.

Networking

The sending of the data from one person to another person is known as **routing** in the internet world.

- There is no need to create VPC, NAC, and any other network interfaces in the serverless computing model as everything related to network is been managed by the AWS, if you need you can create a VPC.
- **CIDR - classless inter-Domain Routing -** It allows you to range between the IP addresses.

VPC

A VPC create a boundary between your AWS architecture and the outside world.

- A VPC protects your content from the outside internet traffic, on your wish.
- The VPC have subnets that are used inside the VPC, as a sub networks.
- We can assign a public or a private internet subnet, but assigning if a subnet is private or public internet cannot be determined by a subnet, it is determined at the route table.

- The CIDR is used in order to assign the range of IP addresses the Subnet or the VPC must use.
- In general, AWs reserves at least 5 IP addresses.
- **Route tables** are used to send the network traffic through out the VPC or into the desired subnets.
- Creating no route tables to a subnet makes it a private subnet, as there is no route table to the subnet it uses the main route table which allows only local traffic to flow through.
- You must secure your VPC's as you allow public network traffic to flow through it, in AWS you have
    - **Network Access Control Lists (ACL's)** - The ACL's are stateless resources, they act as a firewall for the subnets, which allows the inbound and outbound traffic.
    - **Security groups** - the security groups are stateful resources, as they act as a firewall around the instances, in this we can modify them as our wish, we can allow only inbound traffic and deny outbound traffic, which makes them stateful and more secure.

Storage

There are two main types where er can store the data,

1. Block storage - A block storage store a file in the form of many block in it. Making changes to any character inside the file is easier.
2. Object storage - Whereas a object storage stores a file as an entire file. It basically uses the WORM(Write once, Read Many) Model. To make changes to a character in the file we must update the whole file to do it.
3. File storage - This type of storage is mostly used in the file explorer in the windows in which the files are stored in a tree hierarchal manner such that a folder contains a sub-folder and the sub-folder contains your data.

Elastic Block Storage(EBS)

- An EBS is like and external physical disk, where you store you data other than in the disks in your laptop.
- EBS can be linked to any of your services like EC2 and store the data in the EC2 instance into EBS.
- If in any of the case an EC2 instance crashes or if you stop or terminate the instance then your data will be safely present in the EBS.
- You must need to backup your data in the EBS by snapshots.
- You can also connect your EBS with one or more services using the EBS Multi-attach service.
- The maximum EBS storage is 16TB.

Amazon Simple storage service(S3)

- The underlying storage type of S3 is the object type of storage.
- And the each object in S3 can store up to 5 TB individually.
- S3 is a distributed storage, as it is stored among multiple availability zones in the same region.
- And S3 sets the permissions in private by default such that the AWS account user can only see the data in the bucket.
- We can also make policies to the AWS S3 named S3 policies, that are options that can be used in order to make the permissions of who can access the buckets in S3 more granular.
- We can use secure socket layer and client-side encryption in order to encrypt our data when it is in transit.
- There are six type of S3 storage classes. This classes have different properties
    - S3 one-zone - store your data only in one zone
    - S3 intelligent tiering - stores your data in to two tiers by which one tier of data is stored of which is accessible often and the other tier is used for data that is not accessed frequently

    - S3 glacier, glacier deep archine, and other glacier classes are used to store the data that isn't accessed for many days or years
- We can also automate the tier transition by using the transition and the expiration access which are used to change the classes of the data after their creation. for example, if we need to store the data that we created one year before to store into glacier as and archive file as we won't access it frequently we use this method
- Bucket policies are used in order to grant permissions to the users or roles.

Database

The managing of the databases is of 3 types:

1. On-premises - Everything that are needed to manage a database must be done by yourselves, or this might be also called the physical database that are managed on sites.
2. Unmanaged - A few aspects of the the databases are managed by the use rand except other are managed by the AWS.
3. Managed - In this type of database the databases are mostly managed by the AWS, Example AWS RDS.

Amazon RDS

Amazon aurora, is a SQL language that is basically designed for operations in the AWS. It is **5 times** faster from the standard MySQL commands. And the type of this is "Cloud native".

- The AWS RDS are used in order to store the structured data.
- we can control the security of the RDS by placing the RDS in the VPC and the subnets. You can grant permissions such that no one can access the DB, by placing it as a private and also using various ACL's. We can also use the IAM's for the authorization purpose of the database.

- Automatic backups are used by default by AWS RDS, to backup your data. you can assign an automatic backup.
- We create snapshots in the RDS to create manual backups, snapshots are those which have the copy of data and are used while disaster recovery.
- When you backup the data manually it creates an another database instance, which stores the data in it, and the data is used or accessed if any disaster or any data migration or backup compliance occurs.
- We create two AZ’s with a standby DB, and when a problem occurs, in any case if the primary DB fails the standby DB is used an alternative, a same DNS name is provided to the standby DB.

Amazon various databases

- Amazon DynamoDB - it is used to store unstructured data and runs on NoSQL.
- Amazon DocumentDB - it is used to store various types of data that includes both RDS and DyanamoDB.
- Amazon Neptune- it is used for social networking and other recomendation engines, it is a graph database, and is also used for fraud detection.
- Amazon QLDB(Quantum ledger database) - it is an immutable type of database that is used mostly for banking systems and other services which uses services that are not mostly changeable.

Amazon DynamoDB

- The data in the DynamoDB is organized into items and the items contains attributes.
- DynamoDB are very quick is responding, and have less latency and can also handle 10 trillion requests per day.
- Your data is stored into SSDs and replication throughout many AZ’s.
- The items in the table of a DynamoDB are those that contain various attributes and the attributes are nothing but various present in an item.
- the partition key is the primary key of a table.

Monitoring in AWS

- Monitoring the AWS services that we use can be done using the AWS cloud watch.
- The resources that you host your solutions all create various forms of data and usage of the services, each and every service cam be monitored using the metrics which in turn change into statistics.
- Metrics are those which have the data of the utilization of the services and provide a detailed roadmap on how to resolve the problem.
- Each and every service of the AWS create various types of metrices, which indeed are known as the problems that occur or might occur while using the service.
- Monitoring your services can result in, finding the problems that your service have before the end users find it, or stop threat actors and other security concerns that might happen, and also to improve the performance of your resource.
- CloudWatch eases up the monitoring of the data through out your services.
- The AWS Cloud watch can be used to monitor your resources and via their metrices, that records the data of the usage of resources by the services, and we can also set alarms in order to get notified when an metrices goes over the threshold value.
- These alarms are set off when the given threshold value is exceeded and then the notification can be sent to a specific email address, by using the AWS SNS.
- Many AWS services send the metrices per one data point for every 5 minutes, this comes under Basic monitoring.
- The detailed monitoring costs an extra sum than the basic monitoring which is free of cost, the data collection from the metrices is done for every 1 minute in the detailed monitoring.
- Each metric is attached with dimension to vary each and every metric.
- The log events are piled up and combines into a log stream and the log streams are piled up into a log groups.

Elastic Load Balancer

- an ELB(elastic load balancer in the AWS is used in order to manage the traffic between the services and the users.
- The ELB's are designed in a high availability manner, such that there is no chance for the failure of any Elastic load balancers
- The different components of the ALB are target, rule, listener. The listener are the client-side from where the client sends a requests to, a port and a protocol is assigned depending on the load balancer type. and the target is the backend.
- The traffic is sent to the other services using various algorithms and the most popular around in the round robin, which sends the traffic to each server one after another.
- the connection between the ALB and the user is encrypted by using the ACM(Amazon certificate manager) as it provides the SSL certificate or you can import the certificate from the IAM.
- The ALB sends the traffic to each and every server equally using the least outstanding request routing algorithm.
- ALB is specifically used for HTTP and HTTPS.
- Network load balancers can handle millions of request at a second.

EC2 Auto Scaling

Vertical scaling - increasing the size of the instance or different type of instance.

Horizontal scaling- it is done by the EC2 auto scaling based on the metrices using the Cloud watch.

- EC2 auto scaling is a tool that automatically adds or removes the EC2 instance based on the traffic.
- The launch templates in EC2 auto scaling is used in order launch an EC2 instance, you can launch an EC2 instance from scratch or you can launch it by using a template or you can launch it using the properties of the previously launched instance.