**AWS**

In computer terms, a **tool** is a software program or website that has a specific purpose or need. A **service** is a cloud-based or web-based offering that provides resources or functionality to users over the internet.

**Cloud Computing:**

It’s the delivery of different services over the internet, including data storages, servers, databases, networking and software.

Instead of storing files on a storage device or hard drive, a user can save them on cloud, making it possible to access the files from anywhere, as long as they have access to the web.

**The data is stored on physical servers, which are maintained by a cloud service provider [Aws, Azure].**

**Based on Service, cloud computing is categorized into three types:**

* Infrastructure-as-a-service (IaaS)
* Platforms-as-a-service (PaaS)
* Software-as-a-service (SaaS)

**Based on deployment cloud computing is categorized into three types:**

* **Public**
* **Private**
* **Hybrid**

**Elastic Load Balancer: It distributes traffic among multiple virtual servers.**

**Auto-Scaling: Automatic scaling up and scaling down virtual servers.**

**EC2: It is used to create virtual machines known as instance.**

**Ec2 has OS, RAM, Memory, Storage and Networking.**

**Bandwidth: The capacity at which the network can transmit data.**

**Ex: If a bandwidth of a network is 40mbps it implies that the network cannot transmit data faster than 40mbps in any case.**

**Ubiquitous: A service which can be accessed anywhere from the world.**

**AWS services by category:**

1. **Compute**

* **EC2: It is used to create virtual machines known as instance.**
* **Lambda:**
* **Elastic Beanstalk**

1. **Storage**

* **S3**
* **EFS**
* **FSx**
* **S3 Glacier: Archival storage and long-term backups.**
* **Storage Gateway**
* **AWS Backup**
* **AWS Elastic Disaster Recovery: Indian Bank**

1. **Database**

* **RDS**
* **DynamoDB**
* **Elastic Cache**
* **Neptune**
* **Amazon Document DB**

1. **Networking and Content Delivery**

* **VPC**
* **CloudFront**
* **Route53**
* **API Gateway**
* **Direct Connect**
* **AWS Private 5G**

1. **Management and Governance**

* **AWS Organizations**
* **Cloud Watch**
* **AWS Auto Scaling**
* **Cloud Formation**
* **Cloud Trail**
* **AWS Resource Explorer**
* **Control Tower**
* **AWS Chatbot**
* **Launch Wizard**
* **AWS Compute Optimizer**
* **Amazon Grafana**
* **Amazon Prometheus**

1. **Security Identity and Compliance**

* **IAM**
* **Resource Access Manager**
* **Security Hub**

1. **AWS Cost Management**

* **Billing**
* **AWS Cost Explorer**

1. **Containers**

* **Elastic Container Registry**

**It allows you to store Docker images**

* **Elastic Container Service [Container-as-a-service]**

**Allow us to focus on building and deploying our applications without managing the underlying infrastructure.**

* **Elastic Kubernetes Service**

**Open-source container orchestration platform that automates the deployment, scaling, and management of containerized applications.**

**QUESTIONS FOR INTERVIEW**

**1.What is AWS?**

**AWS is a cloud computing platform that offers services such as virtual servers, storage that are used to build and run applications.**

**It allows us to scale resources up and down as and when you need.**

**Drawback: Wide range of services that are difficult to understand for new users.**

1. **Describe three major types of cloud services?**

**Public cloud — data and other information delivered over the internet that can be shared with various people.**

**Largest public cloud providers are Amazon, Azure and GCP.**

**Private cloud — cloud solution that is dedicated to a single organization.**

**Private cloud is best suited for govt agencies.**

**Hybrid cloud — a combination of the two. This environment uses public and private clouds.**

1. **What is IaaS, PaaS, SaaS?**

**IaaS: Aws EC2**

**PaaS: Elastic Beanstalk, EKS [Elastic k8s service], AKS.**

**SaaS: Microsoft Office, G-Mail, Slack.**

1. **How do Availability zone and region relate to one another?**

**An AWS Region is a physical location in the world where we have multiple Availability Zones. Availability Zones are different locations within an AWS Region that are engineered to be isolated from failures in other Availability Zones.**

**Region is a collection of AZ. AZ are data centres of AWS.**

**AZ is nothing but Physical data centre.**

**[20 instances can be created in one availability zone]**

**[At present there are 32 Regions with 102 Availability Zones]**

1. **Describe Auto Scaling.**

**Autoscaling is a cloud computing feature that enables organizations to scale cloud services such as virtual machines up or down automatically based on traffic.**

1. **What is Cloud Front? [Edge Locations]**

**It is a content delivery service provided by Aws it distributes contents such as web pages, images and videos to users across world with low latency.**

**Cloud Front uses a global network of edge locations to deliver the content from the location closest to the end-users, reducing latency and improving the overall performance of your applications.**

1. **What is Amazon Cloud Watch?**

**Amazon cloud Watch logs are stored in S3 and we can use Amazon Elastic Search ? Kibana? to Visualize them.**

1. **What is Instance type in AWS?**

**Instance type is a combination of CPU, memory, storage, and networking capacity it gives us the flexibility to choose the appropriate mix of resources for our applications.**

**[t2.micro is free tier eligible]**

1. **What is Amazon EC2?**

Elastic Compute Cloud is a cloud computing service offered by Amazon; we can deploy our application in Ec2 servers. We can configure the EC2-Instance in a very secure manner by using the VPC, Subnets, and Security group. Based on demand of our application we can scale our instance up and down.

1. **What is Key-pair?**

Key pair is to securely connect to our instance. Key pair is a combination of public key that is used to encrypt data and a private key that is used to decrypt data.

**10.What is Amazon S3? [Simple Storage Service]**

First service provided by AWS is S3.

Amazon **S3 is an object storage service** that stores data as objects within buckets. An object is nothing but a file. A bucket acts as container for objects. S3 bucket name should be unique. S3 has a versioning feature.

Buckets and the objects in S3 are private and can be accessed only if you grant access permissions. You can use bucket policies, AWS Identity and Access Management (IAM) policies, access control lists (ACLs), and S3 Access Points to manage access.

You can use ACLs to grant read and write permissions to authorized users for individual buckets and objects.

When you enable S3 Versioning in a bucket, Amazon S3 generates a unique version ID for each object added to the bucket.

Versioning in Amazon S3 is a means of keeping multiple variants of an object in the same bucket.

There is only one restriction in S3 i.e., individual file cannot be greater than 5TB.

**GET:** Download **PUT:** Upload **List:** To list content of resource.

**Pricing Cost:** Pricing cost in s3 is based on two things

1. Storage Space **ii.** Accessing Cost

**11. What are the storage classes available in amazon S3?**

Depending on the usage we have to choose the classes

**S3 Standard:** It is a default class.

High storage cost, Low Accessing cost.

**S3 Standard-Infrequent Access:** If data is not accessed frequently go with infrequent access.

Low storage cost, High accessing cost.

**Intelligent tier:** If you don’t have any idea like how much often your data being accessed then go with intelligent tier. It will automatically shift your data from one class to other.

[standard to IA, IA to Glacier it depends]

**S3 Glacier:**  It is a low-cost storage service for data archiving and long-term backup.

Archive data consists of older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons.

**12. What is VPC?**

VPC is a crucial component for building and securing AWS resources.

VPC provides an isolated area of the AWS cloud where you can launch AWS resources in a virtual network that we define.

VPC provides us with control over our network architecture, including the ability to create subnets, route traffic and launch resources such as Amazon EC2 instances within your virtual network.

VPC is defined in CIDR block.

**13. What is subnet?**

A subnet is a piece of larger network. **Each subnet is associated with a specific Availability Zone within a region**. Availability Zones are unique, physically separated data centres within a given AWS region. By creating subnets in different Availability Zones, you can achieve high availability and fault tolerance for your applications and services.

**[20 instances can be created in one availability zone]**

**14. Data Archiving?**

Data archiving is the process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons.

**15. What are Security group?**

Security Groups acts like firewall for ec2 instances by controlling inbound and outbound traffic. By default, security groups contain outbound rules that allow all outbound traffic & we can't create rules that deny access. [No deny Only Allow Access]

**16. What is Cloud Formation?**

In simple terms CF is IaaC service. CF template is written in JSON or YAML. We can create any AWS resource by writing Cloud Formation template. Cloud Formation sticks to only AWS. Whereas we can create any cloud infra through terraform.

**17. What is Elastic Beanstalk?**

EBS is a Platform as a service. It will provide platform for developers; developers have to just upload their code and Elastic Beanstalk automatically runs and builds the application.

**18. API? Application Program Interface.**

It establishes communication between two different applications.

**19. Elastic Block Store? EBS**

EBS is **block-storage service designed for Amazon EC2** to store persistent data. This means that the data is kept on the AWS EBS servers even when the EC2 instances are shut down.

**20. What are volumes & snapshots?**

Volume is the **storage area of particular instance** and the **snapshot is the copy of the volume** that can be used as backup data. AWS does not offer a separate backup service so the snapshot acts as a backup service here.

Snapshots are used to create backup for volumes.

A snapshot takes a copy of the EBS volume and places it in Amazon S3, where it is stored redundantly in multiple Availability Zones. **The initial snapshot is a full copy of the volume; ongoing snapshots store incremental block-level changes only**.

*Volume belongs to Availability Zone; Snapshot belongs to region.*

Volume is a storage & snapshot is a backup.

**The data is kept on the AWS EBS servers even when the EC2 instances are shut down.**

**21. What is Elastic file system? [**Elastic = Scalable = Stretch**]**

Fully scalable file storage system. **EFS can increase and decrease its storage capacity depending on the incoming data**. **If you don’t have an idea how much data you are going to store** then you can use the Amazon EFS. You pay only for the storage that you are going to use. While creating EFS it will ask for VPC.

EFS is **Serverless system.**

**22. AWS Lambda? [**serverless compute service**]**

Run code without thinking about servers.

AWS Lambda is sometimes referred to as function-as-a-service.

AWS takes care of the infrastructure provisioning, scaling, and maintenance. You only pay for the compute time your code consumes, and there is no need to worry about managing servers**.**

**23. What is Load Balancer? Types of LB? Elastic LB?**

Rather than giving IP address of individual virtual machine giving the IP address of Load balancer, Load balancer will forward the request to one of the servers.

Distributes traffic evenly across multiple servers.

**[**A load balancer works as a “traffic cop” sitting in front of server and routing client requests across all servers**]**

Load balancer simply distributes the set of requests from clients effectively across multiple servers and ensures that no single server is busy with too many requests that lead to downtime of application.

**Application load balancer** operates on Layer 7 of OSI model it knows http requests.

**Network load balancer** operates on layer 4 transport layer.

**Classic LB** (not in use).

Load balancer distributes incoming traffic among group of servers that increases the availability of application.

**24. What is a cluster? [Collection of servers - cluster]**

A cluster is a group of servers that runs as if it is a single entity.

**25. IAM? Components of IAM?**

**Authentication:** It is a process of verifying a user or a device before allowing access to a system or resource.

**Authorization:** It is the process of giving users the ability to access a resource.

**26. Multi factor authentication**

It is a multi-step account login process that requires users to enter more information than just a password. For example, along with the password, users might be asked to enter a code sent to their email, answer a secret question, or scan a fingerprint.

**27. IAM Control list?**

Create and manage IAM user IAM group, Manage the security credentials of user, Create and manage the policies to grant access to AWS services.

**28. IAM role and IAM user?**

**29. What is Amazon Route 53. [Domain Name System]**

Route 53 on AWS provides DNS as a service.

Route 53 will map the IP address with domain name.

**30. Latency?**

Time taken by a data to transfer across the network.

Networks with a longer delay have high latency, while those with fast response times have low latency.

**31.**

**32. AWS config? AWS Cloud Trail?**

Aws configure is assessing AWS from command line. It will ask for Access key ID and Secret access key.

Cloud Trail enables us to **monitor and record activities and events within our AWS infrastructure**.

**33. Amazon EC2:**

Ec2 is a service provided by Amazon where we can rent for virtual servers, virtual servers known as instances.

Key features of EC2:

* **Scalability:** Users can easily scale the number of instances up or down based on their computing needs.
* **Instance type:** EC2 offers a wide range of instance types with different combinations of CPU, memory, storage, and networking capabilities, allowing users to choose the most suitable instance type for their specific workloads.
* **Operating System Support:** EC2 instances support various operating systems, including various flavours of Linux, Windows, and others. Users can choose the operating system that best fits their application requirements.
* **Load Balancing:** AWS provides load balancing services like Elastic Load Balancing (ELB) that can distribute incoming traffic across multiple EC2 instances to improve availability and fault tolerance.
* **Auto Scaling:** Users can configure Auto Scaling to automatically adjust the number of running EC2 instances based on traffic.

**34. Which type of scaling you prefer for RDS? Vertical or Horizontal?**

**35. Amazon EC2 instance types.**

General Purpose Instances.

Compute Optimized Instances.

Memory-Optimized Instances.

Accelerated Computing Instances.

Storage Optimized Instances.

**36. Difference between spot instance, reserved instance and**

**on-demand instance.**

* **On-demand instances:** Whenever we require an instance will go for it, this instance is pay-as-you-go with **no long-term commitments**. Whenever we don’t need, we just terminate the instance.
* **Reserved instance:** Instead of requesting as & when required we reserve this instance for **either 1 or 3 years**. **Cost is low** when compared to On-demand instance.
* **Spot instance:** A Spot Instance is an unused EC2 instance that is available for less than the On-Demand price Will get **90% discount** compared to On-demand instance. Amazon EC2 will give you a two-minute warning when AWS needs the capacity back**.**

**37. In AWS three different storage services are available. When to use EFS, S3, elastic?**

**38. Edge Locations:** These are content delivery network [CDN’s] endpoints, located in major cities around the world specifically used by cloud front to distribute content to end user to reduce latency. [400+ Edge Locations. 102-AZ]

All Edge Locations have copy of actual data.

[Latency: Time taken by a data to transfer across the network]

**39.** **Local Zone: [**Kolkata & Delhi**]** It is an extension of an Aws region, resources created in local zones can serve local users with low latency. We can add local zones as an availability zone in a region.

**41. Wavelength Zone:** Wave length zone can produce mobile applications with low latency.

**42.** **ARN: Amazon Resource Name:** Unique name given by amazon to each resource.

Example: arn:aws:s3:::*bucket\_name/key\_name*

ARN is used to give specific permissions for a resource to a particular user.

**43. Public IP: [**It allows resources to communicate with internet**]**

Allows resources, such as Amazon EC2 instances or load balancers, to communicate directly with the internet.

**44. Private IP: [**For internal communication within VPC**]**

In AWS, a private IP address is an address that is used within a Virtual Private Cloud (VPC) and is not directly accessible from the public internet. Private IP addresses are assigned to resources such as Amazon EC2 instances, RDS databases, and other services within a VPC. These addresses are used for internal communication and to create isolated networks within AWS.

**If a private network wants to access internet, then NAT gateway is must.**

**Terms:**

* **Server:** A computer that provides services or resources or data toanother computer.
* **Shell Scripting:** Sequence/List of commands which we execute at once.

**Private IP Ranges:**

* 10.0.0.0/8 to 10.255.255.255
* 172.16.0.0/12 to 172.31.255.255
* 192.168.0.0/16 to 192.168.255.255
* Highly Available: Operates continuously without failing.

DOCKER INSTRUCTIONS:

1. **FROM** This is usually the first instruction in a Docker file and specifies the base image on which your image will be built.
2. **RUN** Used to install packages.
3. **COPY** It is used to copy your local directories to Docker Container.
4. **ADD** Add is used to copy files from local directory as well as from urls.
5. **WORKDIR** It sets Current working directory.
6. **EXPOSE**  Container will listen on specific port.
7. **CMD**
8. **ENTRYPOINT**
9. **USER**  To create a user inside container.

A **SERVER** is a software or hardware device that accepts and responds to requests made over a network. The device that makes the request, and receives a response from the server, is called a client. On the Internet, the term "server" commonly refers to the computer system that receives requests for a web file and sends those files to the client.

REGION **→**AVAILABILITY ZONE**→** DATA CENTRE **→**RACK **→**BLADESERVERS**→** EC2

General Purpose Instances.

Compute Optimized Instances.

Memory-Optimized Instances.

Storage Optimized Instances.

Accelerated Computing Instances.