**AWS Solution Architect (Associate)**

* **What is Cloud Computing?**

Cloud Computing is on demand delivery of compute power, database, storage, applications and other IT resources through a cloud Service platform via the internet with pay-as-you-go paying model.

* **AWS stands for?**

Amazon Web Services.

* **Top players of Cloud Services**

1. AWS
2. Microsoft Azure
3. GCP (Google Cloud Platform)
4. Alibaba Cloud Services
5. Oracle
6. VMWare Cloud Services
7. Fujitsu cloud services
8. IBM

* **Characteristics of Cloud**

1. **On Demand self-service:**

On-demand self-service is a characteristic of cloud computing that allows users to provision, manage, and control computing resources as needed without requiring human intervention from the service provider. Essentially, it enables users to access and deploy resources such as virtual machines, storage, and applications through automated processes via the internet.

1. **Broad Network Access:**

Broad network access is another fundamental characteristic of cloud computing. It refers to the ability for users to access cloud services and resources over the internet or other wide area networks using a variety of devices, such as laptops, smartphones, tablets, or desktop computers. This accessibility is typically provided through standard internet protocols and allows users to connect to cloud services from anywhere with an internet connection.

1. **Scalability:**  
   Scalability in the context of computing refers to the ability of a system to handle growing amounts of work or to accommodate an increasing number of users without sacrificing performance or reliability. In cloud computing, scalability is a crucial feature that allows resources to be easily adjusted to meet changing demands.

**There are two main types of scalabilities:**

1. **Vertical Scalability:** Also known as scaling up, this involves increasing the capacity of individual components within a system, such as upgrading a server's CPU, memory, or storage capacity.
2. **Horizontal Scalability:** Also known as scaling out, this involves adding more instances of resources, such as additional servers, to distribute the workload across multiple machines. This approach enables systems to handle larger loads by adding more resources in parallel.
3. **Resource Pooling:**  
   Resource pooling is a fundamental concept in cloud computing that involves the aggregation of computing resources, such as servers, storage, and networking devices, into a shared pool. This pooled infrastructure is then dynamically allocated and reallocated to users on an as-needed basis.

**Key aspects of resource pooling include:**

1. **Multi-tenancy:** Multiple users or tenants share the same physical resources, with each user's data and applications logically isolated from one another for security and privacy.
2. **Dynamic Provisioning:** Resources are allocated and reallocated dynamically based on demand, allowing for flexible and efficient utilization of available capacity.
3. **Elasticity:** The pool of resources can scale up or down dynamically to accommodate fluctuations in demand, ensuring that users have access to the necessary computing resources when needed.
4. **Measure Services:**

"Measured services" is a term used in cloud computing to describe the capability of cloud service providers to monitor and track resource usage accurately. Essentially, it involves the measurement, monitoring, and reporting of resource consumption by users or applications in the cloud environment.

Overall, measured services provide transparency, accountability, and flexibility in cloud computing by enabling customers to monitor and manage their resource usage effectively while ensuring accurate billing and cost management.

**Here are a few key points about measured services:**

1. **Usage Tracking:** Cloud providers track the consumption of resources such as computing power, storage, bandwidth, and other services used by customers. This tracking enables accurate billing based on actual usage.
2. **Metering:** Resource usage is measured and quantified using predefined metrics, such as the amount of data stored, the number of compute hours used, or the volume of network traffic generated.
3. **Billing and Cost Management:** Measured services play a crucial role in determining the cost of using cloud services. Customers are billed based on their actual usage, allowing for transparent and pay-as-you-go pricing models.
4. **Performance Optimization:** By analysing usage data, customers can gain insights into their resource utilization patterns and identify opportunities for optimizing performance and cost-efficiency.

* **AWS History**

Launch in 2006. And got improved.

* **AWS Certification**

1. AWS Solution Architect (Associate and Professional)
2. AWS Devops (Developer + Operations) (Associate and Professional)
3. AWS Sysops (System + Operations) (Associate and Professional)

* **Services Model in Cloud:**

1. **IAAS (Infrastructure as a Service)**

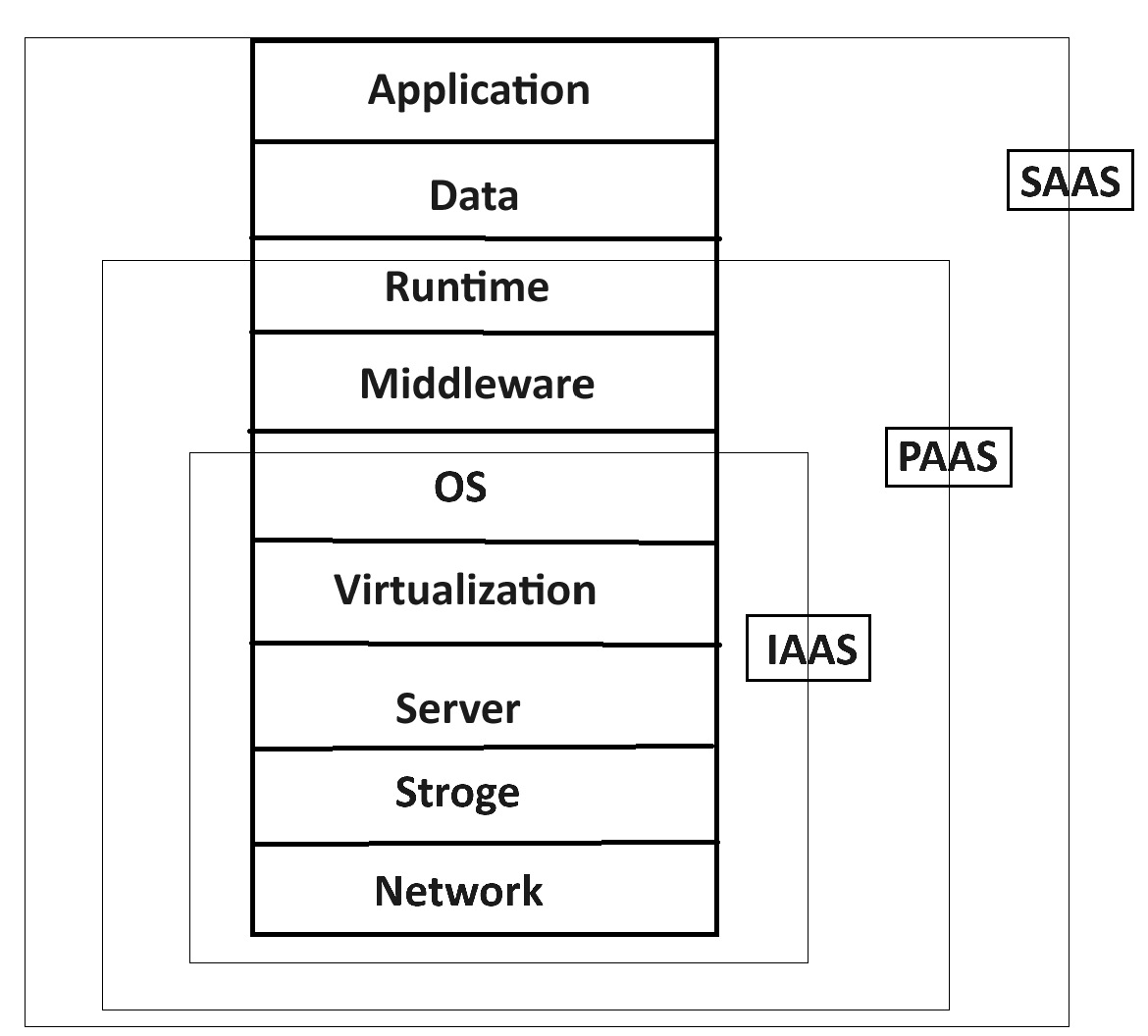
This includes Network, Storage, Server, Virtualization and Operating system. Users have control over the Network, Storage, Server, Virtualization Operating system and development frameworks running on the provided infrastructure.

1. **PASS (Platform as a service)**

PaaS provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the underlying infrastructure. PaaS offerings typically include development tools, middleware, databases, and other components required for application development and deployment.

1. **SAAS (Software as a service)**

SaaS delivers software applications over the internet on a subscription basis, eliminating the need for users to install, maintain, and manage the software locally. Applications are accessed through web browsers or APIs, and the provider handles maintenance, updates, and security. Common examples of SaaS include email services like Gmail, productivity suites like Microsoft Office 365, and customer relationship management (CRM) software like Salesforce.



* **Deployment Model of Cloud**

Deployment Model of Cloud describes how cloud computing resources are provisioned, managed, and accessed. Each deployment model offers unique advantages and challenges

**There are 3 primary deployment models:**

1. **Public Cloud:** In a public cloud deployment model, cloud services and infrastructure are provided by a third-party cloud service provider and are available to the general public over the internet. Resources such as virtual machines, storage, and applications are shared among multiple users, and customers typically pay for the resources they consume on a pay-as-you-go basis. Public clouds offer scalability, flexibility, and cost-effectiveness, making them popular among startups, small businesses, and enterprises. Examples of public cloud providers include **Amazon Web Services (AWS)**, **Microsoft Azure**, and **Google Cloud Platform (GCP)**.
2. **Private Cloud (Enterprise Cloud):** **A private cloud deployment model involves the provisioning and management of cloud resources within a dedicated infrastructure that is owned and operated by a single organization**. Unlike public clouds, private clouds are not shared with other organizations, providing greater control, security, and customization options. Private clouds can be hosted on-premises or by third-party providers, and they are often used by enterprises with specific security, compliance, or performance requirements. Examples of private cloud solutions include **VMware Cloud Foundation**, **OpenStack**, and **Microsoft Azure Stack**.
3. **Hybrid Cloud:** A hybrid cloud deployment model combines elements of both public and private clouds, allowing organizations to leverage the benefits of both environments. In a hybrid cloud setup, workloads can be dynamically distributed between on-premises infrastructure and public cloud services based on factors such as cost, performance, and compliance requirements. This flexibility enables organizations to optimize resource utilization, scalability, and resilience while maintaining control over sensitive data and applications. Hybrid cloud architectures are increasingly common among enterprises with diverse IT environments and evolving business needs.

* **Hypervisor**

A hypervisor, also known as a virtual machine monitor (VMM), is software or firmware that creates and manages virtual machines (VMs) on a physical host machine. It allows multiple operating systems (OS) to run simultaneously on a single physical hardware platform.

Different cloud service providers may use different hypervisors to power their virtualization infrastructure.

AWS -> Nitro Hypervisor/Citrix

Azure -> Hyper-V

**Note:** To make cloud we must need to virtualize the infra then only we can make cloud.

* **Elastic Compute Cloud (EC2)**

1. EC2 provides scalable computing capacity in the AWS cloud. Using Amazon EC2 eliminates your need to invest in hardware up front so we can develop deploy applications faster.
2. Amazon EC2 enables you to scale up or down to handle changes in requirements or spikes in popularity reducing your need to forecast traffic.
3. We can use Amazon EC2 to launch as many or as few virtual servers as we need configure security and networking and manage storage.
4. It enables us to scale up or scale down the instances.
5. It having two storage options **EBS** and instance store (all will be erased like RAM)
6. Preconfigured templates are available known as Amazon Machine Images
7. By default, when you create an EC2 account with Amazon your account is limited to a maximum of 20 instances per EC2 Region with two default high I/O instances.

* Type of EC2 Instances

1. General Purpose
2. Compute Optimized
3. Memory Optimized
4. Storage Optimized
5. Accelerated Computing/GPU
6. High Memory
7. Previous Generation