Cloud Computing

What is Cloud Computing?

Cloud computing is the on-demand delivery of compute power, database storage, applications, and other IT resources through a cloud services platform with pay-as-you-go pricing. It provides:

- Provisioning of exactly the right type and size of computing resources.
- Access to as many resources as needed, almost instantly.
- A simple way to access servers, storage, databases, and a set of application services.
- Amazon Web Services (AWS) owns and maintains the network-connected hardware, while you provision and use what you need via a web application.

The Deployment Models of the Cloud

Private Cloud	Public Cloud	Hybrid Cloud
Cloud services used by a single organization, not exposed to the public.	Cloud resources owned and operated by a third-party cloud service provider, delivered over the Internet.	Keep some servers on-premises and extend some capabilities to the cloud.
Complete control over data, security, and compliance.	Cost-effective as infrastructure is shared among multiple users.	Allows data and applications to be shared between private and public clouds.
Security for sensitive applications, ideal for critical workloads.	Suitable for less sensitive workloads that require high scalability and availability.	Offers flexibility, security, and scalability for different use cases.
Meet specific business needs and compliance requirements.	No maintenance required as the cloud provider manages the infrastructure.	Provides business continuity, disaster recovery, and data backup solutions.

The Five Characteristics of Cloud Computing

1. **On-demand self-service**: Provision computing resources as needed automatically.
2. **Broad network access**: Access cloud resources over the network using standard mechanisms.
3. **Resource pooling**: Providers serve multiple customers using a multi-tenant model.
4. **Rapid elasticity**: Resources can be scaled up or down rapidly.
5. **Measured service**: Resource usage is monitored and billed accordingly.

Six Advantages of Cloud Computing

1. **Cost Savings**: Pay only for the computing power, storage, and other resources you use.
2. **Speed and Agility**: Quickly deploy services and resources.

3. **Scalability**: Easily scale resources up or down as needed.
4. **High Availability**: Highly available architecture for business continuity.
5. **Global Reach**: Access services from any geographical region.
6. **Security**: AWS provides robust security capabilities to protect your data.

Problems Solved by the Cloud

- **High upfront costs**: Replaced by a pay-as-you-go model.
- **Scalability limitations**: Dynamic scaling to meet business demands.
- **Limited infrastructure availability**: Global infrastructure to support workloads.

Types of Cloud Computing

Infrastructure as a Service (IaaS)	Platform as a Service (PaaS)	Software as a Service (SaaS)
Provides virtualized computing resources over the internet (e.g., AWS EC2). Offers maximum control over the infrastructure. Suitable for developers needing control over OS, middleware, and runtime.	Provides a platform allowing customers to develop, run, and manage applications (e.g., AWS Elastic Beanstalk). Focus on deploying applications without managing underlying infrastructure. Ideal for developers who want to focus on application development.	Provides software applications over the internet on a subscription basis (e.g., AWS Chime). Accessible over the internet, usually via a web browser. Suitable for users needing access to software without infrastructure management.

Example of Cloud Computing Types

- **IaaS**: AWS EC2 (Elastic Compute Cloud)
  - GCP, Azure, Rackspace, Digital Ocean, Linode
- **PaaS**: AWS Elastic Beanstalk
  - Heroku, Google App Engine (GCP), Windows Azure (Microsoft)
- **SaaS**: AWS Chime
  - Google Apps (Gmail), Dropbox, Zoom

Pricing of the Cloud − Quick Overview

AWS follows three fundamental pricing principles based on the pay-as-you-go pricing model:

Fundamental	Description
Compute	Pay for the compute time you consume. Examples include EC2 instance hours or Lambda invocation duration.
Storage	Pay for the amount of data stored in the cloud. Examples include S3 storage space and EBS volume usage.
Data Transfer OUT	Pay for data transfer out of the cloud. Data transfer IN is free. This pricing structure solves the issue of expensive data transfer fees in traditional IT systems.

How Cloud Pricing Solves Traditional IT Cost Issues

- Traditional IT requires expensive upfront investments for hardware, maintenance, and upgrades.
- With AWS's pay-as-you-go model, you only pay for what you use, reducing overall costs.
- You can scale up or down based on demand, minimizing under-utilized resources.

AWS Cloud Use Cases

1. **Web Hosting**: Host websites with elastic scaling and high availability.
2. **Big Data Analytics**: Run analytics on large datasets.
3. **Application Hosting**: Host applications with global accessibility and automated scaling.
4. **Disaster Recovery**: Implement disaster recovery strategies with minimized infrastructure.
5. **Backup and Storage**: Store backups in a highly durable and secure manner.

AWS Global Infrastructure

AWS Regions

- Geographically isolated areas where AWS clusters data centers.
- Each region has multiple Availability Zones.
- Used to deploy applications close to customers for lower latency.

How to Choose an AWS Region?

- **Latency**: Choose a region closest to your customers for lower latency.
- **Compliance**: Ensure the region meets data residency and compliance requirements.
- **Services Available**: Check which AWS services are offered in the region.
- **Pricing**: Prices vary by region, so choose a region that fits your cost requirements.

AWS Availability Zones (AZs)

- Multiple, isolated data centers within a region.
- Each AZ has independent power, cooling, and networking.
- Provides redundancy and fault tolerance in case of a failure.
- They're connected with high bandwidth, ultra-low latency networking

AWS Points of Presence (Edge Locations)

- Network locations that deliver content closer to end users.
- Used by services like Amazon CloudFront and AWS Global Accelerator.
- Provides low latency and improved performance for content delivery.

AWS Shared Responsibility Model

What is the Shared Responsibility Model?

- AWS and the customer share responsibility for security and compliance.
- Divides security tasks based on **AWS as the provider** and **customer as the user** of cloud services.

AWS Responsibilities: Security *of* the Cloud

- AWS is responsible for protecting the infrastructure that runs all services offered in the AWS Cloud.
- Includes hardware, software, networking, and facilities:
  - **Physical security** of data centers (e.g., access control, environmental controls).
  - **Infrastructure** security, such as maintaining hypervisors, host operating systems, and network infrastructure.
  - **Global network** operations, such as DDoS protection and monitoring.

Customer Responsibilities: Security *in* the Cloud

- Customers are responsible for managing and securing what they put in the cloud.
- Includes:
  - **Data protection**: Encrypt data in transit and at rest.
  - **IAM**: Control access through Identity and Access Management (IAM) roles, users, and policies.
  - **OS and application configurations**: Maintain security of guest operating systems, applications, and firewall configurations.
  - **Network settings**: Manage security group rules and network access control lists (NACLs).
  - **Compliance**: Ensure compliance with regulations and standards based on data storage and usage.

Example Responsibilities for Different AWS Services

Service Type	AWS Responsibility	Customer Responsibility
IaaS (e.g., EC2)	Securing physical infrastructure, hypervisor, and network.	Configure and secure OS, patch management, data, and network settings.
PaaS (e.g., RDS)	Managing the database engine, backups, and patching.	Secure data at rest and in transit, manage DB access, and IAM roles.
SaaS (e.g., S3)	Protecting the service's underlying infrastructure.	Manage permissions, bucket policies, and data lifecycle rules.

Shared Responsibility Model

Figure 1: Shared Responsibility Model

Summary

- AWS handles security *of* the cloud, while customers manage security *in* the cloud.
- Understanding your responsibilities helps you implement appropriate security measures for your AWS environment.