

AWS Solution Architect (UNIT-1: AWS Cloud Foundations, AWS IAM)

Dr. Munsifa Firdaus Khan Barbhuyan Assistant Professor-Grade(II)

School of Computing Science and Engineering & Artificial Intelligence Email: munsifafirdauskhan@vitbhopal.ac.in

UNIT-1



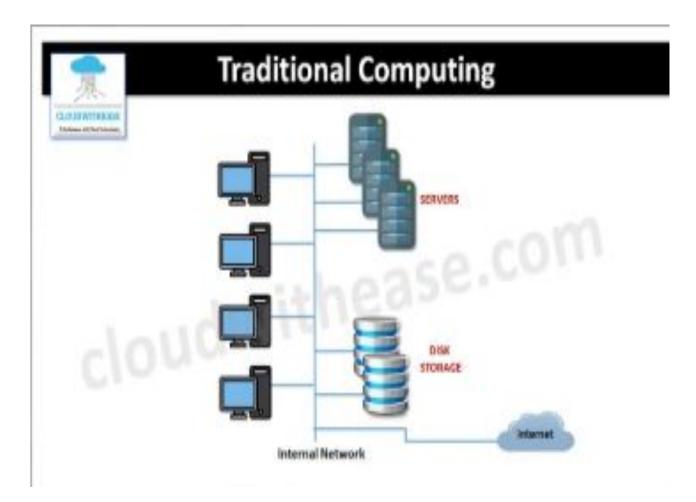
- AWS Cloud Foundations: Introduction to Cloud Computing, Overview of AWS Global Infrastructure, AWS Shared Responsibility Model
- AWS IAM AWS Identity and Access Management (IAM), IAM Users, Groups, and Policies, IAM Roles, IAM Best Practices
- LAB 1: Creation of IAM User & Policies.
- LAB 2: Creation of IAM Groups.
- LAB 3: Creation of IAM role and assign to IAM user with policies.

What is Traditional Computing?





- Traditional computing refers to the use of physical hardware like personal computers, servers, and storage devices for running applications and processing data.
- In traditional computing, all the resources, software, and data are typically stored and managed locally on physical machines that an organization or individual owns or leases.



What is Traditional Computing?





Key Characteristics:

- On-Premise Infrastructure.
- Fixed Resources.
- Maintenance and Management
- Upfront Capital Investment
- Limited Scalability
- Fixed Location and Access
- Security and Control
- In-House Support

What is Traditional Computing?



Real-World Example:

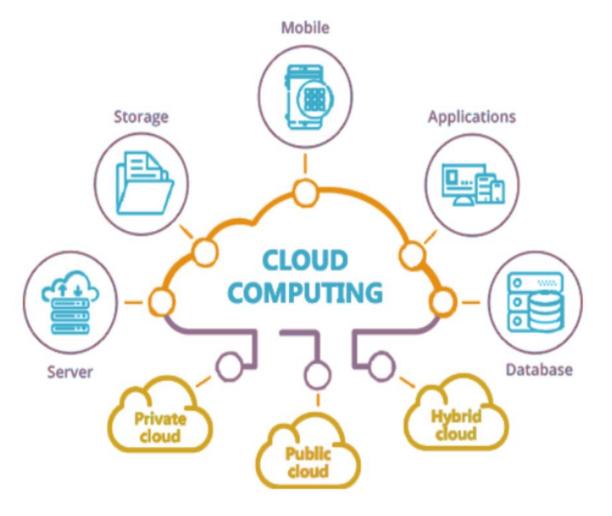
• Corporate Data Centers: Large companies like Banks and Financial Institutions (e.g., JPMorgan Chase, Bank of America), Government Agencies (e.g., U.S. Department of Defense, NASA), Media and Broadcast Companies (e.g., BBC, NBC) and Healthcare Providers (e.g., Kaiser Permanente, Cleveland Clinic).

• **Personal Computers:** A typical user with a desktop or laptop computer that stores files and runs applications locally is another example of traditional computing.

What is Cloud Computing?

BHOPAL www.vitbhopal.ac.in

- Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing.
- Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on as-needed basis from a cloud provider like Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), IBM Cloud, Oracle Cloud, Salesforce etc.



What is Cloud Computing?



Key Characteristics:

- On-Demand Access: Cloud services are available over the internet, and users can scale resources up or down based on their needs (eg: AWS, Azure, or Google Cloud).
- Pay-As-You-Go: Cloud computing generally uses a subscription or usage-based pricing model, which reduces upfront costs (eg: AWS EC2 instances).
- Scalable and Flexible: Resources like storage, processing power, and software can be scaled up or down quickly without physical hardware limitations.
- Managed Services: The cloud provider handles maintenance, updates, and security.
- **Resource Pooling**: Cloud providers pool computing resources to serve multiple customers, dynamically allocating and reassigned resources according to demand (eg: A cloud provider may run multiple virtual machines (VMs) for different users on the same physical hardware but ensure each user gets isolated and secure resources.).

What is Cloud Computing?

Key Characteristics:



- Automation: Cloud computing allows for the automation of various processes like provisioning, scaling, backup, and system updates, reducing the need for manual intervention (eg: AWS Lambda enables serverless computing, where code can be executed automatically based on triggers without needing manual server management).
- Multi-Tenancy: Cloud environments are shared by multiple customers (tenants), with the cloud service provider ensuring data separation and privacy between them (eg: A SaaS provider like Salesforce hosts several customers on the same infrastructure but ensures data and configurations are isolated and secure).
- Security and Privacy: Cloud providers invest heavily in security technologies to protect their infrastructure, networks, and data. Many offer encryption, identity management, and compliance with regulatory standards (eg: AWS Identity and Access Management (IAM) and encryption)
- **Broad Network Access**:Cloud services are available over the internet and can be accessed from anywhere using standard devices like laptops, smartphones, or tablets (eg: **Google Drive**).

Who is using cloud computing?



- Organizations of every type, size, and industry are using the cloud for a wide variety of use cases, such as data backup, disaster recovery, email, virtual desktops, software development and testing, big data analytics, and customer-facing web applications.
- For example, healthcare companies are using the cloud to develop more personalized treatments for patients. Financial services companies are using the cloud to power real-time fraud detection and prevention. And video game makers are using the cloud to deliver online games to millions of players around the world.
- Netflix, Dropbox, Spotify, Airbnb etc.

Traditional Computing VS Cloud Computing





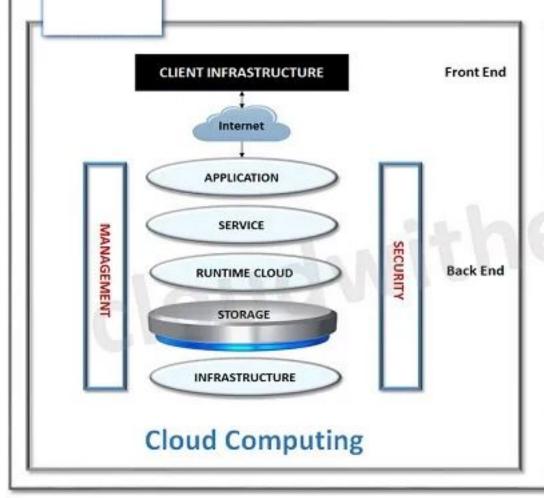
Function Cloud Computing		Traditional Computing	
Business model	Pay for use (subscription model), administrative overhead is reduced	Pay for assets, administrative costs	
Concept	Delivery of different services such as data and applications though internet on different servers	·	
Data access	Ability to access data anywhere at any time by end user	User can access data only on system in which it is stored	
Costs	Most cost effective as operations and maintenance of server is shared among several parties which reduces cost of public services	Less cost effective as one has to buy expensive equipment to operate and maintain server	
Connectivity	Requires fast, reliable, and stable Internet connection to access data and application	Does not require Internet connectivity to access data or application	
Scalability and Elasticity	It is highly scalable and elastic one can increase or decrease computing resources as per business need		
Resiliency and redundancy	Resiliency and redundancy are built in by cloud providers	Resiliency and redundancy levels depend on architecture, additional costs are involved to built a high resiliency architecture	

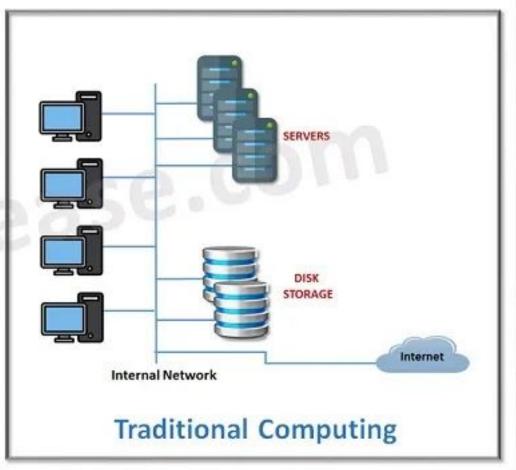
Traditional Computing VS Cloud Computing





Cloud Computing vs Traditional Computing







The three main types of cloud computing include Infrastructure as a Service, Platform as a Service, and Software as a Service. Each type of cloud computing provides different levels of control, flexibility, and management so that you can select the right set of services for your needs.

Infrastructure as a Service (IaaS)

- laaS contains the basic building blocks for cloud IT. It typically provides access to networking features, computers (virtual or on dedicated hardware), and data storage space. laaS gives you the highest level of flexibility and management control over your IT resources. It is most similar to the existing IT resources with which many IT departments and developers are familiar.
- AWS offers services like EC2 (Elastic Compute Cloud) for running virtual machines, S3 (Simple Storage Service) for scalable storage, and VPC (Virtual Private Cloud) for networking.



Platform as a Service (PaaS)

- PaaS removes the need for you to manage underlying infrastructure (usually hardware and operating systems), and allows you to focus on the deployment and management of your applications. This helps you be more efficient as you don't need to worry about resource procurement, capacity planning, software maintenance, patching, or any of the other undifferentiated heavy lifting involved in running your application.
- Google App Engine (GAE) is a fully managed platform for building and deploying applications. It automatically handles scaling, load balancing, and application monitoring. Developers can use it to build apps in languages such as Java, Python, PHP, and Go

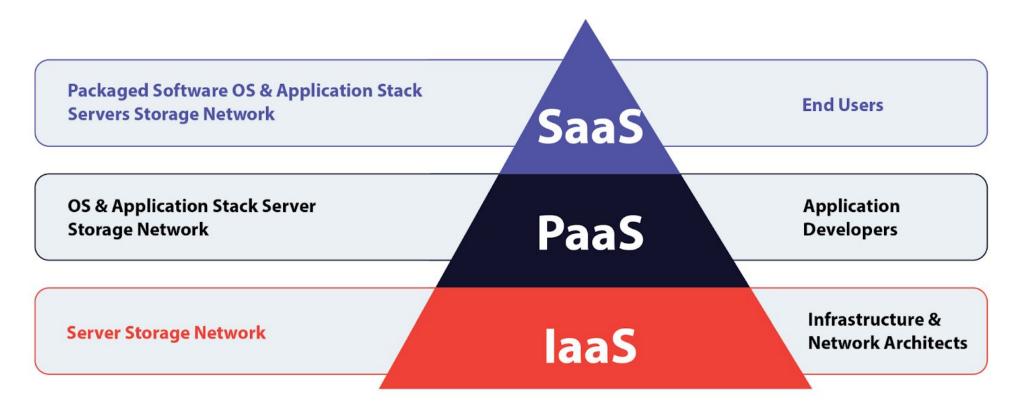


Software as a Service (SaaS)

- SaaS provides you with a complete product that is run and managed by the service provider. In most cases, people referring to SaaS are referring to end-user applications (such as web-based email). With a SaaS offering, you don't have to think about how the service is maintained or how the underlying infrastructure is managed. You only need to think about how you will use that particular software.
- Google Workspace provides cloud-based productivity tools like Gmail, Google Docs,
 Google Sheets, Google Drive, and Google Meet. These tools allow for real-time collaboration and are accessible from any device with internet access



Cloud Service Models



UNIVERSITY OF THE UNIVERSITY O



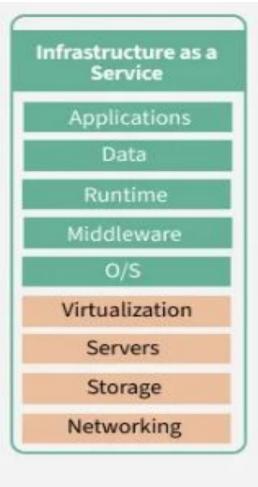
Difference between IaaS, PaaS, and SaaS

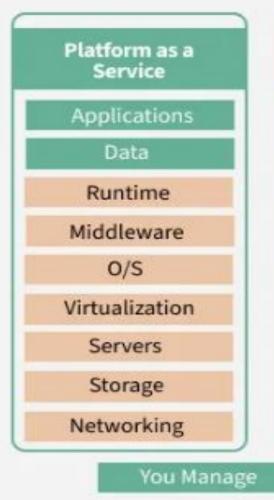
laaS	Paas	SaaS
It provides a virtual data center to store information and create platforms for app development, testing, and deployment.	It provides virtual platforms and tools to create, test, and deploy apps.	It provides web software and apps to complete business tasks.
It provides access to resources such as virtual machines, virtual storage, etc.	It provides runtime environments and deployment tools for applications.	It provides software as a service to the end-users.
It is used by network architects.	It is used by developers.	It is used by end users.
laaS provides only Infrastructure.	PaaS provides Infrastructure+Platform.	SaaS provides Infrastructure+Platform +Software.

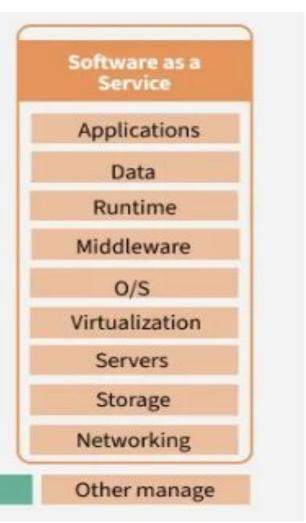
BHOPA www.vitbhopal.ac.in

Difference between IaaS, PaaS, and SaaS







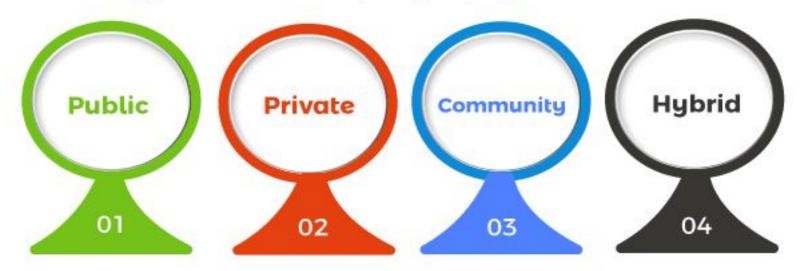


Cloud Deployment Model



• It works as your virtual computing environment with a choice of deployment model depending on how much data you want to store and who has access to the Infrastructure.

Types of Cloud Computing Deployment Models



Cloud Deployment Model





Important Factors to Consider	Public	Private	Community	Hybrid
Setup and ease of use	Easy	Requires professional IT Team	Requires professional IT Team	Requires professional IT Team
Data Security and Privacy	Low	High	Very High	High
Scalability and flexibility	High	High	Fixed requirements	High
Cost- Effectiveness	Most affordable	Most expensive	Cost is distributed among members	Cheaper than private but more expensive than public
Reliability	Low	High	Higher	High

What are cloud services?



- Cloud services are IT resources managed by AWS and delivered on demand over the internet. Traditionally, organizations had to purchase and configure everything from server hardware and storage systems to networking and security technologies before launching any digital system. Provisioning and managing IT infrastructure is expensive, complicated; and takes time away from innovation.
- Cloud services allow anyone to access the IT infrastructure needed to build and maintain digital systems, abstracting complex infrastructure so anyone can build sophisticated applications quickly and scale globally. Running your applications on a cloud server is just the start. You can use cloud services to add artificial intelligence and machine learning (AI/ML), real-time data analytics, and many other capabilities to your applications

What are cloud managed services?



- Cloud services are also called cloud managed services because the underlying infrastructure is fully managed by AWS. All required hardware, operating systems, and other infrastructure layers are stored and managed in highly secure AWS data centers distributed around the globe. We purchase and maintain all types of IT resources, making them available as services you can access in your application code.
- AWS monitors and maintains cloud servers, storage, networks, databases, and more, ensuring consistent performance and uptime. Beyond hardware maintenance, we handle all types of IT operational tasks from load balancing and patch management to disaster recovery and more.
- Today, thanks to AWS, anyone—from college students to enterprise teams—can access cloud services at a fraction of the cost of managing on-premises infrastructure. Anyone can build and deploy software without heavy upfront IT infrastructure investments

Introduction to AWS (Amazon Web Services)





Amazon Web Services (AWS) is one of the world's leading cloud computing platforms, providing a wide array of on-demand computing resources, tools, and services over the internet. Launched in 2006 by Amazon, AWS has become a pivotal enabler for organizations of all sizes to innovate and scale efficiently without the need for substantial upfront capital investment.

Key Features of AWS:

- Scalability: AWS provides elastic computing capabilities, allowing users to scale resources up or down based on demand, ensuring cost efficiency.
- Global Infrastructure: AWS operates in multiple regions worldwide, each with multiple availability zones, ensuring high availability and low latency.
- Comprehensive Services: AWS offers over 200 fully featured services, including compute (Amazon EC2), storage (Amazon S3), databases (Amazon RDS), machine learning (Amazon SageMaker), and more.
- Pay-as-You-Go Pricing: AWS follows a usage-based pricing model, enabling businesses to pay only for the services they use.
- **Security:** AWS prioritizes security with features like encryption, robust access controls, compliance certifications, and threat detection tools.





Benefits of Using AWS:

- Cost Efficiency: AWS eliminates the need for traditional on-premises hardware and maintenance, reducing operational costs.
- Innovation: AWS's extensive suite of tools and services accelerates development cycles, fostering innovation.
- Flexibility: AWS supports a wide range of technologies, frameworks, and programming languages, catering to diverse application needs.
- **Resilience:** High reliability and disaster recovery mechanisms ensure minimal downtime and data protection.

The Beginnings of AWS: A Brief History



1. Founding and Early Days of AWS (2000–2006)

- Amazon, initially (early 2000s) focused on e-commerce, developed internal tools to manage its computing resources efficiently, enabling teams to operate independently while scaling its infrastructure to meet rapid growth.
- By 2003, Amazon realized that the infrastructure it developed for internal use could be turned into a service for external customers. This led to the idea of offering "Amazon's Web Services" to the public, allowing other companies to use Amazon's infrastructure to build and run their own applications.
- In 2006, AWS officially launched its first two services: **Amazon S3 (Simple Storage Service)** for scalable object storage and **Amazon EC2 (Elastic Compute Cloud)** for flexible, on-demand computing capacity. These services laid the foundation for what would become the world's most comprehensive and widely adopted cloud platform.

2. AWS's Early Success and Growth (2006–2010)

- This early success of Amazon S3 and EC2 encouraged AWS to expand its service offerings beyond storage and compute. Over the next few years, AWS introduced services for databases (Amazon RDS), content delivery (Amazon CloudFront), and security (AWS IAM).
- By 2010, AWS had already established itself as a major player in cloud computing, with a rapidly growing customer base that included startups, enterprises, and government organizations.

The Beginnings of AWS: A Brief History Cont.

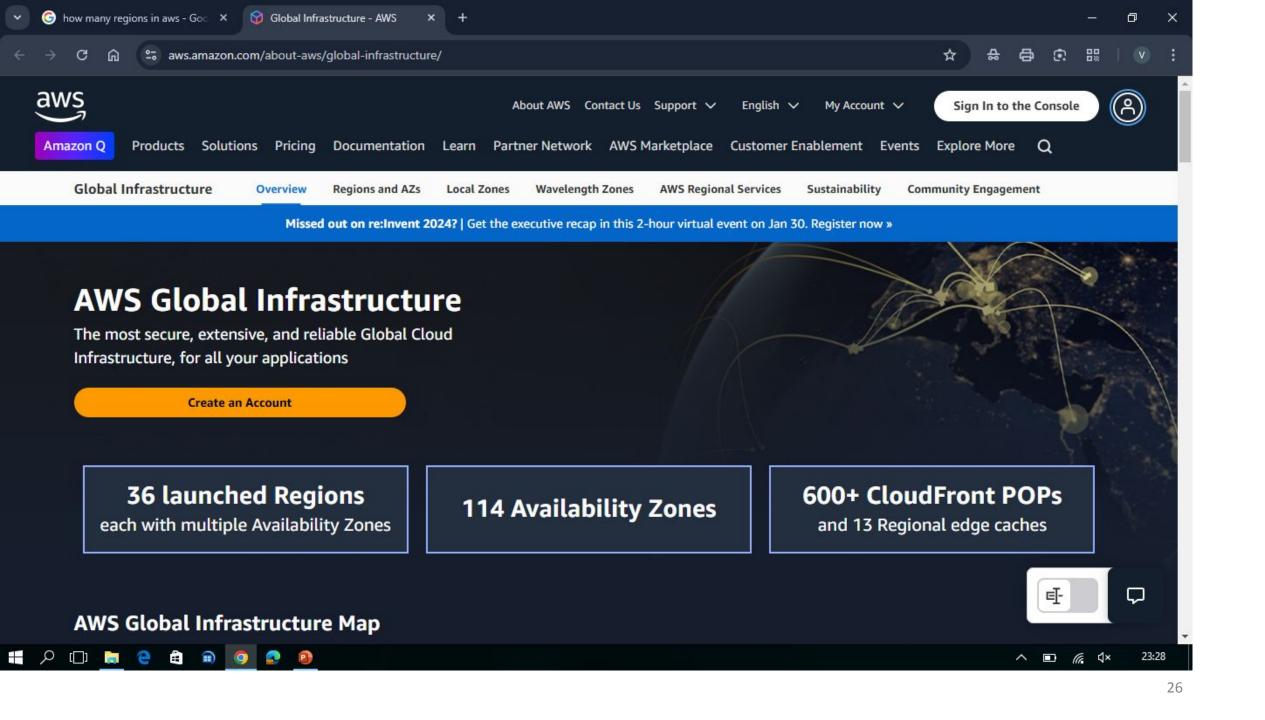


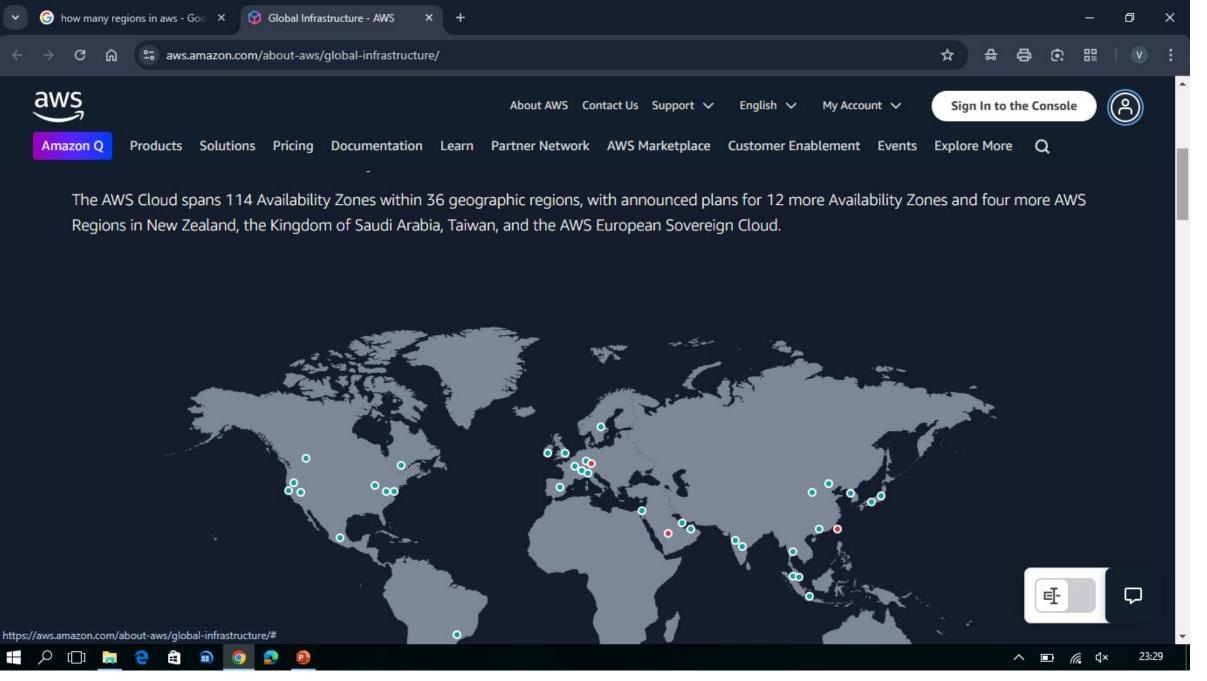
3. Rapid Expansion of AWS Services (2010–2015)

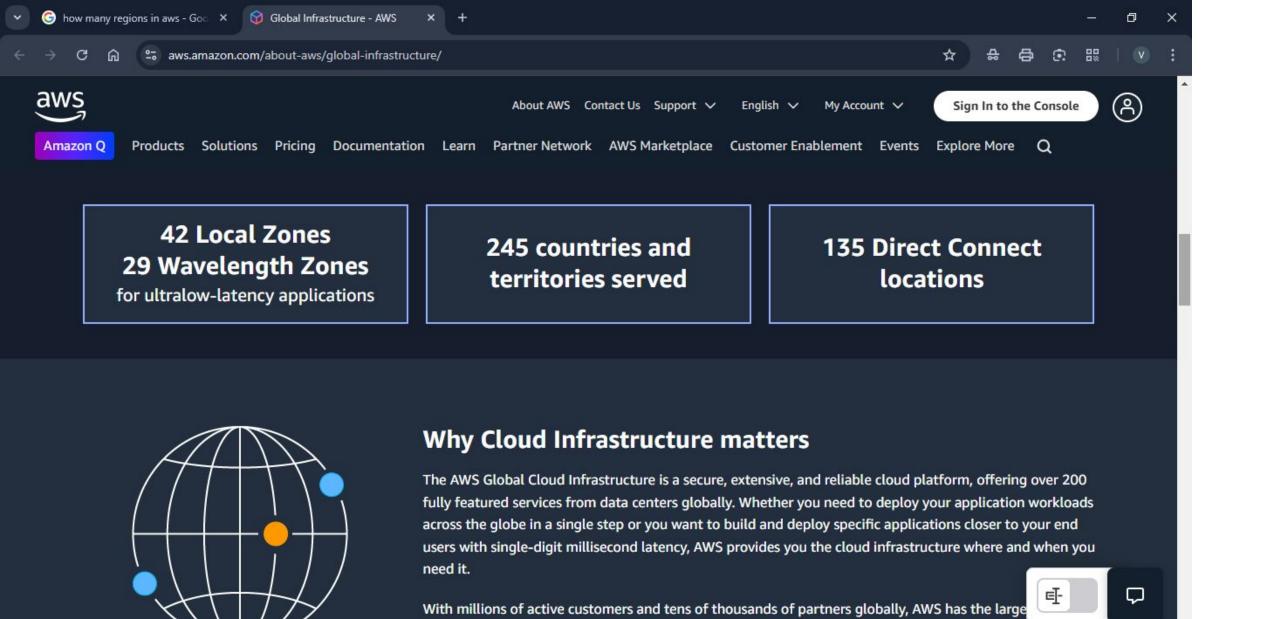
- During this time, AWS rapidly expanded its offerings, launching services like **Amazon RDS** for managed databases, **Amazon Redshift** for data warehousing, and **AWS Lambda** for serverless computing, broadening its support for use cases such as data analytics and machine learning.
- AWS expanded globally by establishing data centers in various regions, enhancing application performance for end-users. It also launched the AWS Marketplace, allowing third-party vendors to offer complementary software and services.

4. AWS's Dominance and Market Leadership (2015-Present)

- By 2015, AWS led the cloud computing market, introducing innovations like **Amazon SageMaker** for machine learning, **AWS Lambda** for serverless computing, and IoT services, alongside strategic acquisitions and partnerships to strengthen its offerings.
- Today, AWS leads cloud computing, offering 200+ services globally. Its innovation and extensive ecosystem make it the preferred platform for businesses of all sizes.







dynamic ecosystem. Customers across virtually every industry and of every size, including start-ups,

23:29

^ ■ @ Q×





- AWS is a cloud computing platform which is globally available
- Global infrastructure is a region around the world in which AWS is based. Global infrastructure is a bunch of high-level IT services.

The following are the components that make up the AWS infrastructure:-

- Availability Zones
- Region
- Edge Locations
- Regional Edge Caches

Components of Global Infrastructure

Availability Zones

Region

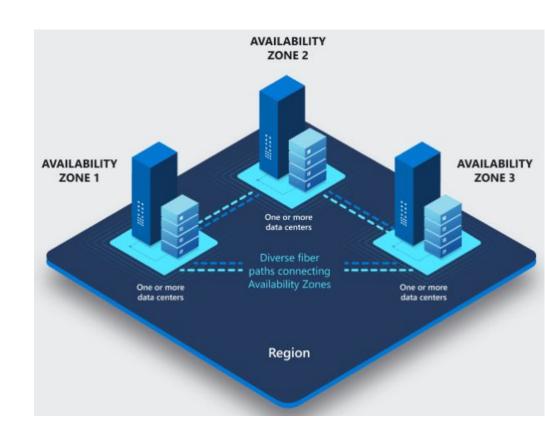
Edge Locations Regional Edge Caches





Availability Zone as a Data Center

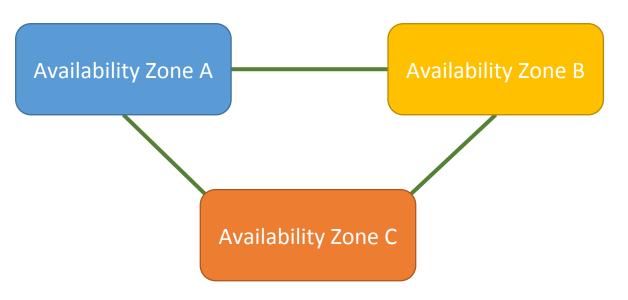
- An availability zone is a facility that can be somewhere in a country or in a city. Inside this facility, i.e, the Data Centre, we can have multiple servers, switches, load balancing, firewalls. The things that interact with the cloud sit inside the data centers.
- A data center can also deliver cached content to your global end-users to improve response times. At its core, the AWS Global Infrastructure utilizes multiple data centers.
- An availability zone can be several data centers, but if they are close together, they are counted as 1







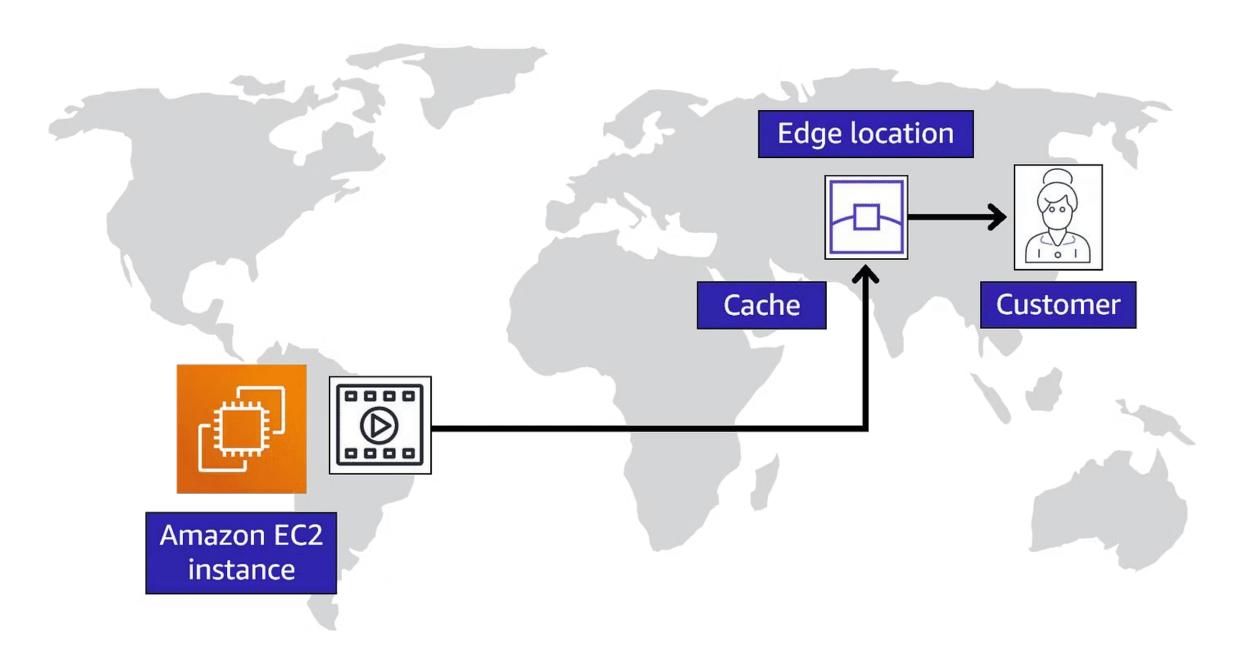
- A region is a geographical area. Each region consists of 2 more availability zones.
- A region is a collection of data centers which are completely isolated from other regions.
- A region consists of more than two availability zones connected to each other through links.

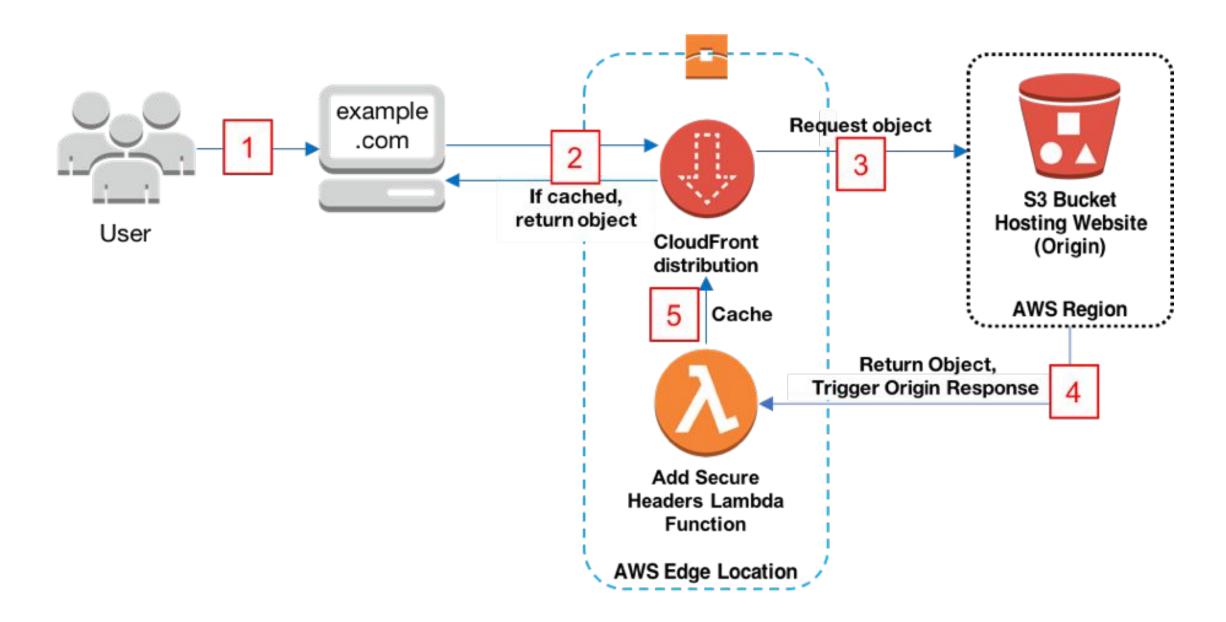




Edge Locations

- Edge locations are the endpoints for AWS used for caching content.
- Edge locations consist of CloudFront, Amazon's Content Delivery Network (CDN).
- Edge locations more than regions. Currently, there are over 150 edge locations.
- Edge location is not a region but a small location that AWS have It is content.
- Edge locations are mainly located in most of the major cities to distribute the content to end users with reduced latency.
- For example, some user accesses your website from Singapore; then this request would be redirected to the edge location closest to Singapore where cached data can bel read.

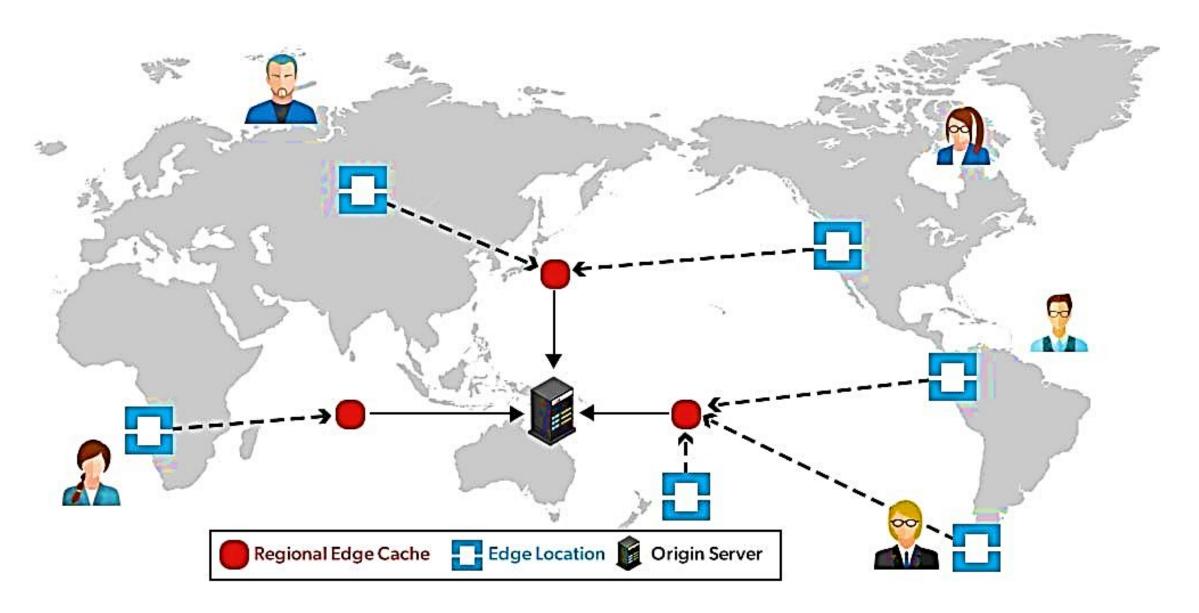


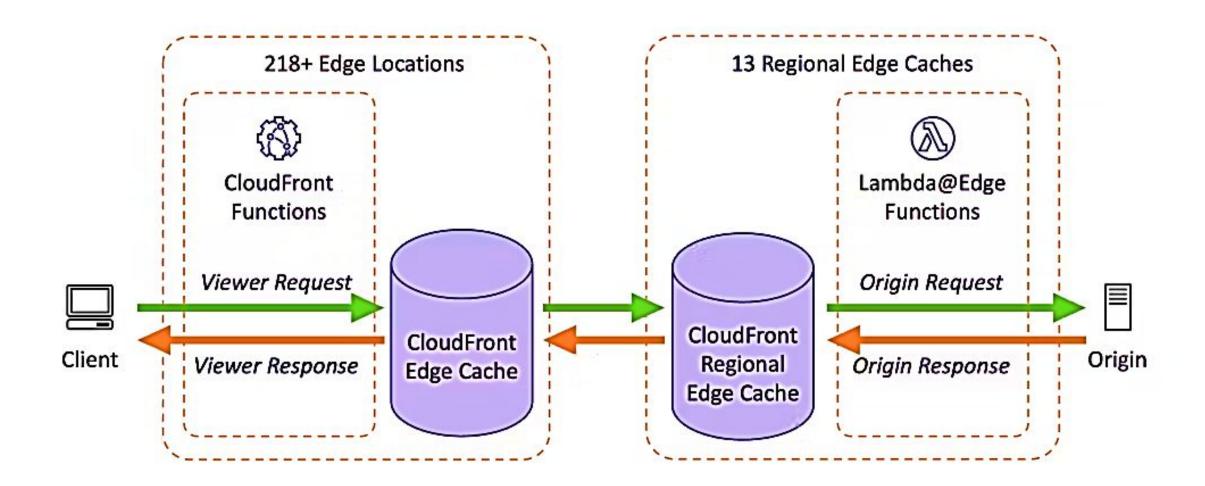


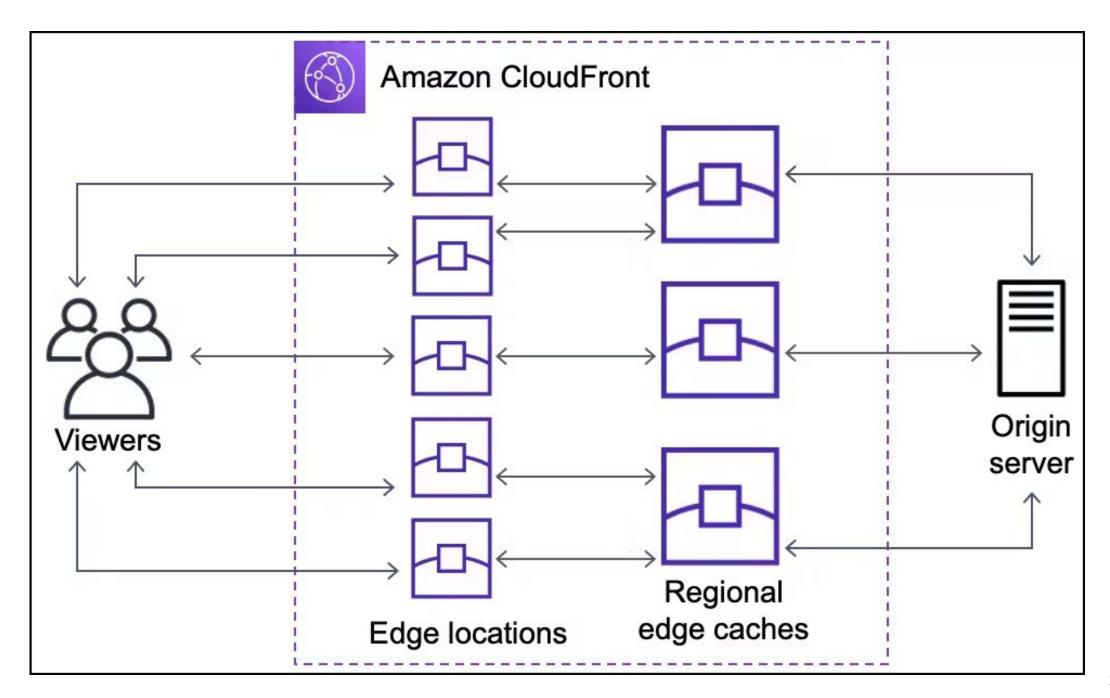


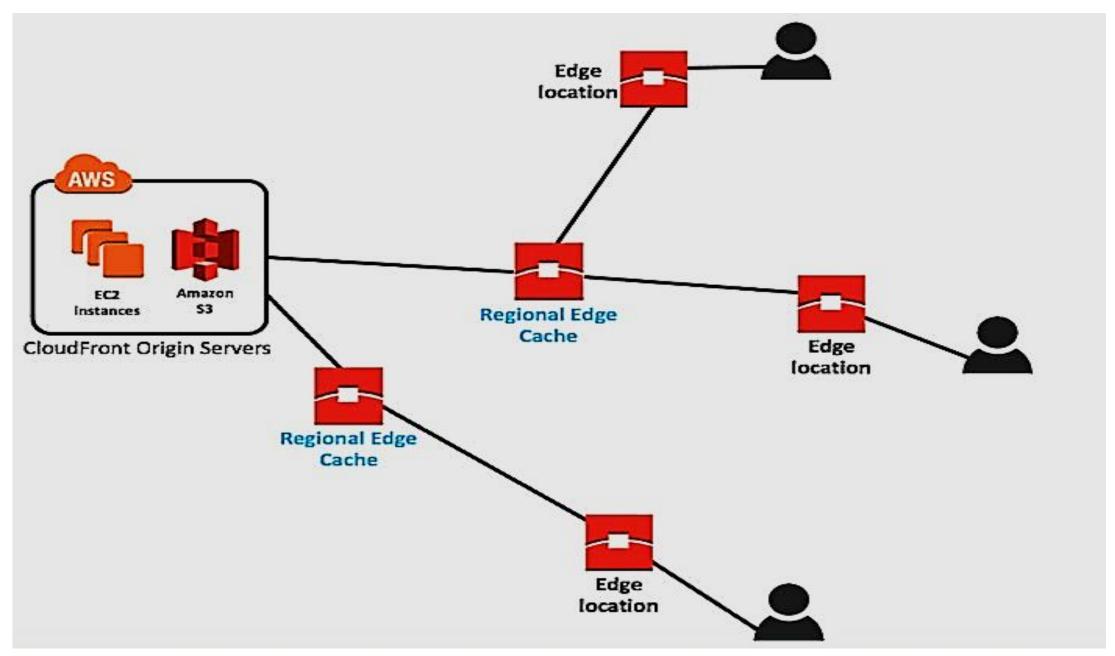
Regional Edge Cache

- AWS announced a new type of edge location in November 2016, known as Regional Edge Cache Edge.
- Regional Edge Cache lies between CloudFront Origin servers and the edge locations.
- A regional edge cache has a large cache than individual edge location.
- Data is removed from the cache at the edge location while the data is retained at the Regional Edge Caches.
- When the user requests the data, then data is no longer available at the edge location. Therefore, the edge location retrieves the cached data from the Regional Edge Cache instead of the origin servers that have high latency.









Amazon Global Infrastructure Cont.



Concepts of Zone, Region and Multi-Region

- **Regions** are independent areas that consist of zones. They affect the pricing, reliability, networking, and performance of zonal resources VM.
- A **zone** is a deployment area for Google Cloud resources within a region. The zone should be considered a single failure domain within a region. To deploy fault-tolerant applications with high availability and help protect against unexpected failures, deploy your applications across multiple zones in a region.
- Multi-region services are designed to be able to function following the loss of a single region. Multi-regional resources are cloud storage, Big Query, Big Tables, etc.



The AWS Shared Responsibility Model outlines how security and compliance responsibilities are divided between AWS and its customers.

- AWS is responsible for "security of the cloud," encompassing the infrastructure
- While customers are responsible for "security in the cloud," including their data, applications, and access controls within their AWS environment.

2. AWS Responsibilities ("Security of the Cloud"):

- **Protecting the infrastructure:** This includes the physical facilities, hardware, software, and networking that run AWS Cloud services.
- Managing access to infrastructure: AWS controls who can access the underlying infrastructure and resources.
- Providing a secure platform: AWS ensures the cloud environment is secure and reliable.



Customer Responsibilities ("Security in the Cloud"):

- Managing their own data and applications: Customers are responsible for the security of their data, applications, and the configurations of their AWS services.
- Implementing access controls: Customers manage who can access their resources and data within their AWS environment.
- Configuring and patching their software: Customers are responsible for managing the security of their guest operating systems, applications, and other software components.

3. Shared Responsibilities:

- **Controls:** Some controls are shared between AWS and the customer. For example, AWS provides the infrastructure for patching, while the customer must implement the patching process for their applications.
- **Compliance:** Both AWS and the customer are responsible for ensuring compliance with relevant regulations and standards.

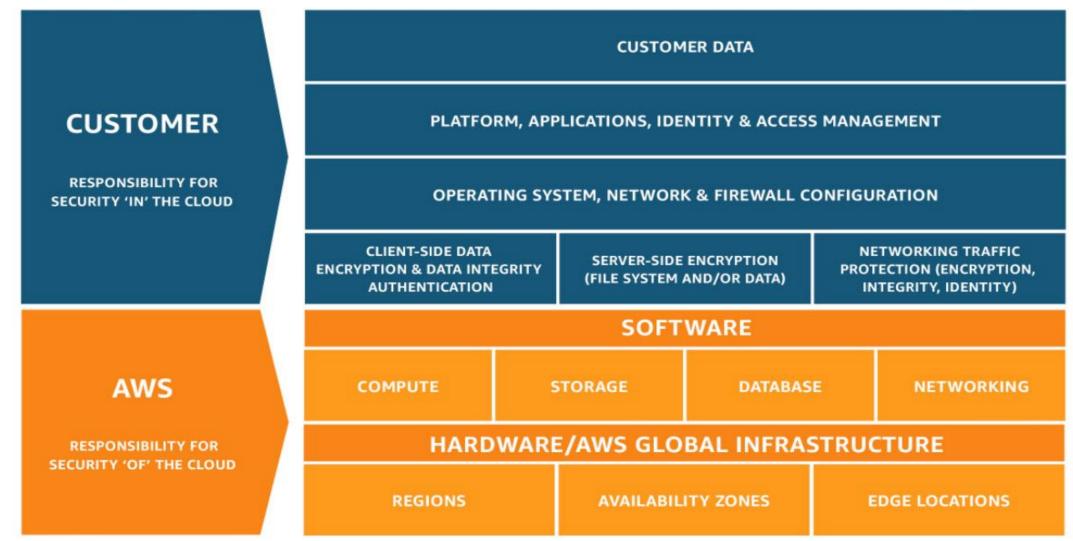


Examples:

- 1. Physical security: AWS is responsible for the physical security of the data centers, while the customer is responsible for the security of their on-premises data centers hosting AWS Outposts.
- **2. Operating system:** AWS manages the security of the operating system for certain services, while customers are responsible for security configurations and patching of their chosen operating systems.
- **3. Data:** Customers are responsible for encrypting their data at rest and in transit, managing access to their data, and ensuring the confidentiality and integrity of their data.

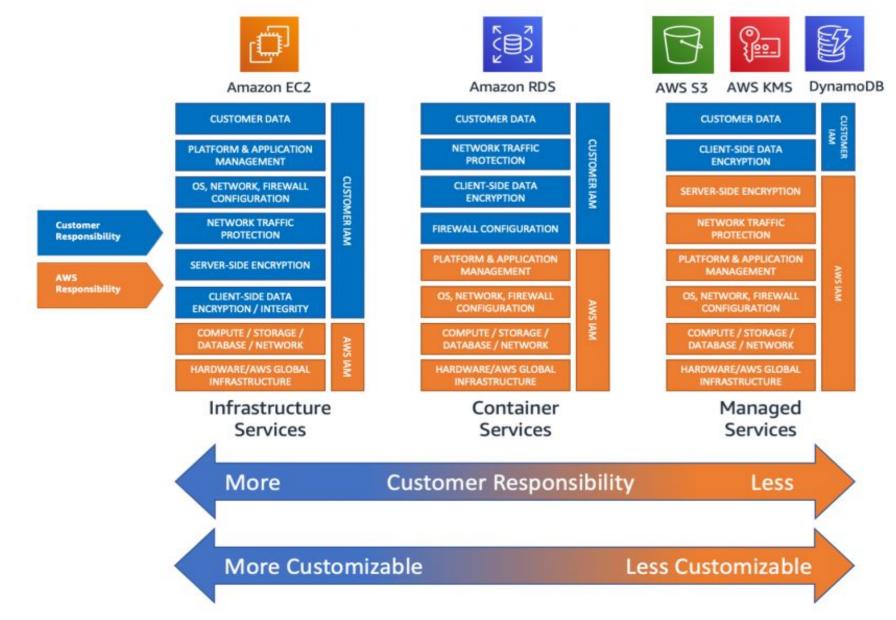












What is AWS IAM?







- IAM stands for Identity Access Management.
- IAM allows you to manage users and their level of access to the AWS console.
- It is used to set users, permissions and roles. It allows you to grant access to the different parts of the aws platform.
- AWS Identity and Access Management is a web service that enables Amazon Web Services (AWS) customers to manage users and user permissions in AWS.
- With IAM, Organizations can centrally manage users, security credentials such as access keys, and permissions that control which AWS resources users can access.
- Without IAM, Organizations with multiple users must either create multiple user accounts, each with its own billing and subscriptions to AWS products or share an account with a single security credential. Without IAM, you also don't have control about the tasks that the users can do.
- IAM enables the organization to create multiple users, each with its own security credentials, controlled and billed to a single AWS account. IAM allows the user to do only what they need to do as a part of the user's job.





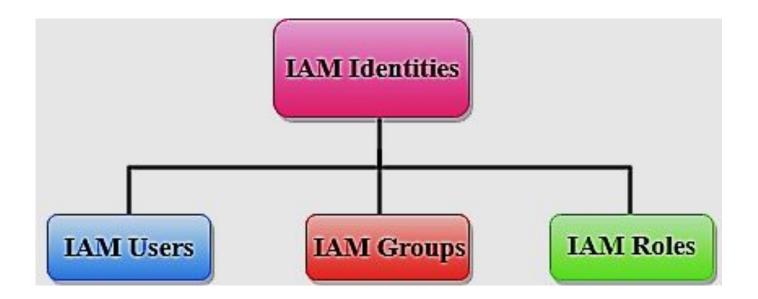
The IAM workflow includes the following six elements:

- A principal is an entity that can perform actions on an AWS resource. A user, a role or an application can be a principal.
- Authentication is the process of confirming the identity of the principal trying to access an AWS product. The principal must provide its credentials or required keys for authentication.
- Request: A principal sends a request to AWS specifying the action and which resource should perform it.
- Authorization: By default, all resources are denied. IAM authorizes a request only if all parts of the request are allowed by a matching policy. After authenticating and authorizing the request, AWS approves the action.
- Actions are used to view, create, edit or delete a resource.
- Resources: A set of actions can be performed on a resource related to your AWS account.





• IAM identities are created to provide authentication for people and processes in your aws account.





IAM Users

- An IAM user is an identity with an associated credential and permissions attached to it. This could be an actual person who is a user, or it could be an application that is a user.
- With IAM, you can securely manage access to AWS services by creating an IAM user name for each employee in your organization.
- Each IAM user is associated with only one AWS account. By default, a newly created user is not authorized to perform any action in AWS.
- The advantage of having one-to-one user specification is that you can individually assign permissions to each user.



IAM Groups

- A collection of IAM users is an IAM group.
- You can use IAM groups to specify permissions for multiple users so that any permissions applied to the group are applied to the individual users in that group as well.
- Managing groups is quite easy. You set permissions for the group, and those permissions are automatically applied to all the users in the group.
- If you add another user to the group, the new user will automatically inherit all the policies and the permissions already assigned to that group. This lessens the administrative burden.



IAM Policies

- An IAM policy sets permission and controls access to AWS resources.
- Policies are stored in AWS as JSON documents. Permissions specify who has access to the resources and what actions they can perform.
- For example, a policy could allow an IAM user to access one of the buckets in Amazon S3.
 The policy would contain the following information:
 - Who can access it
 - What actions that user can take
 - Which AWS resources that user can access
 - When they can be accessed



IAM Policies

```
In JSON
format:
  "Version": "2017-10-17",
  "Id": "S3-Account-Permissions",
  "Statement": [{
                                                       Specify Actions(Read/Write/Delete)
  "Sid": "AddPublicReadPermissions",
  "Effect": "Allow",
                                                       Give permissions(Allow/Deny)
  "Principal":"*",
                                                       Who can Access it
  "Action": "s3:*",
                                                       What action can a user take
  "Resource": ["arn:AWS:s3:::bucket/*"
                                                        Specify the resource
```



IAM Policies

There are two types of policies: managed policies and inline policies.

- A managed policy is a default policy that you attach to multiple entities (users, groups, and roles) in your AWS account. Managed policies, whether they are AWS-managed or customer-managed, are stand-alone identity-based policies attached to multiple users and/or groups.
- Inline policies are policies that you create that are embedded directly into a single entity (user, group or role).





IAM Roles

- An IAM role is a set of permissions that define what actions are allowed and denied by an entity in the AWS console. It is similar to a user in that it can be accessed by any type of entity (an individual or AWS service). Role permissions are temporary credentials.
- For example, you might want to allow a mobile app to use AWS resources, but you do not want it to save the key, credential or password. Or you might want to give access to resources to a user who already has an identity defined outside of AWS, such as a user who already has Google or Facebook authentication. If you want to provide someone with a service or let someone access resources in your account, you can use roles for that purpose too. You also might want to grant temporary access to your account to a third party, such as a consultant or an auditor. They're not permanent users, just users with temporary access to your environment.



IAM Features

- Shared Access to your Account: A team working on a project can easily share resources with the help of the shared access feature.
- Free of cost: IAM feature of the Aws account is free to use & charges are added only when you access other Amazon web services using IAM users.
- Have Centralized control over your Aws account: Any new creation of users, groups, or any form of cancellation that takes place in the Aws account is controlled by you, and you have control over what & how data can be accessed by the user.
- Grant permission to the user: As the root account holds administrative rights, the user will be granted permission to access certain services by IAM.
- Multifactor Authentication: Additional layer of security is implemented on your account by a third party, a six-digit number that you have to put along with your password when you log into your accounts.