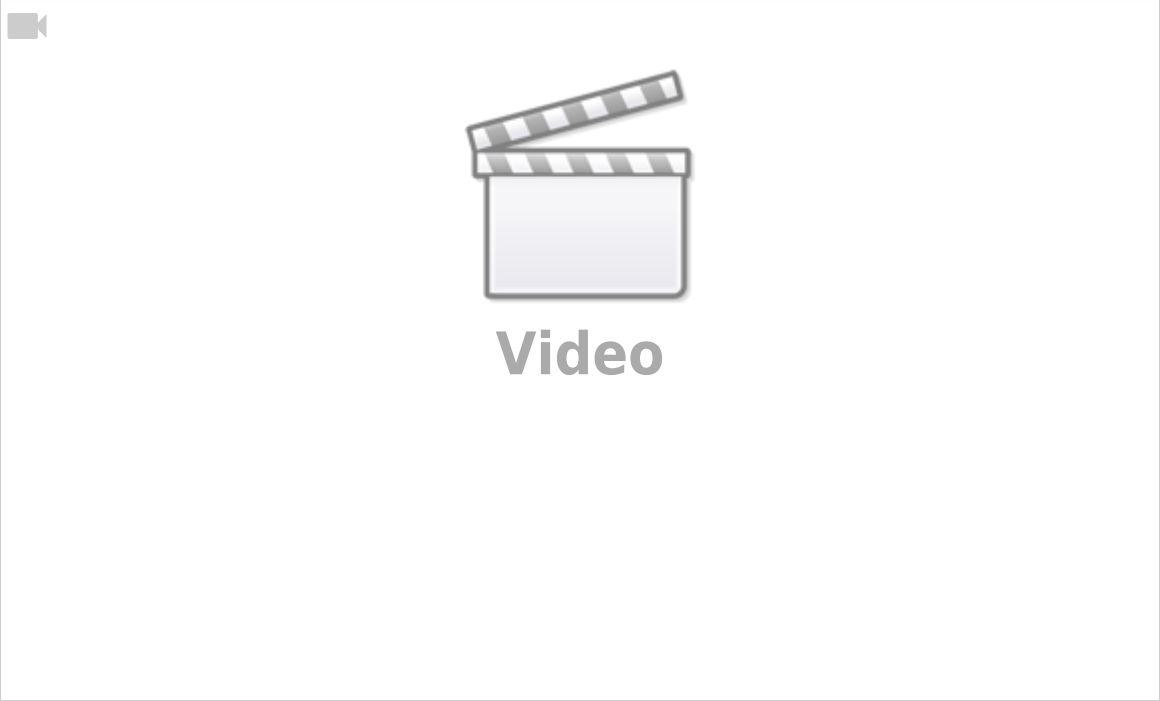
Model Deployment and Cloud for ML

Cloud Computing Essentials

Create and Activate AWS Account



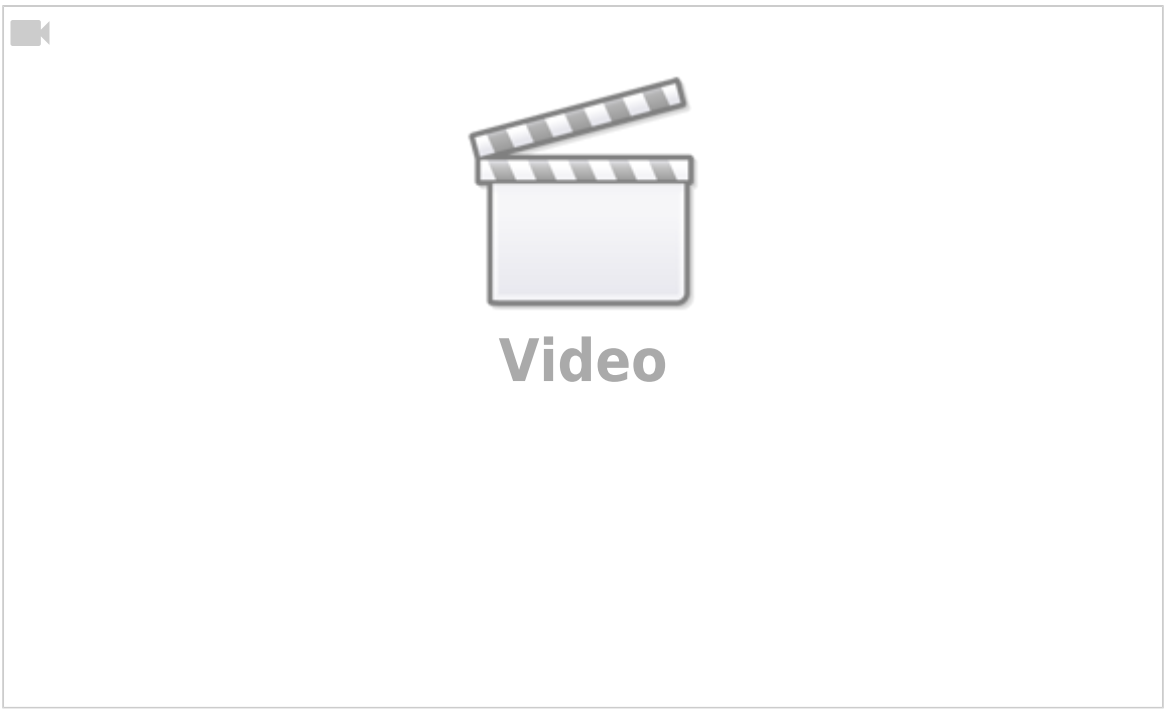
Cloud Computing Basics

Introduction to Cloud Computing

What is Cloud Computing?

Cloud computing, often referred to as simply the “cloud“, means storing and accessing data and programs over the Internet rather than the hard drive of your computer. The data can be anything such as files, images, documents, and more.

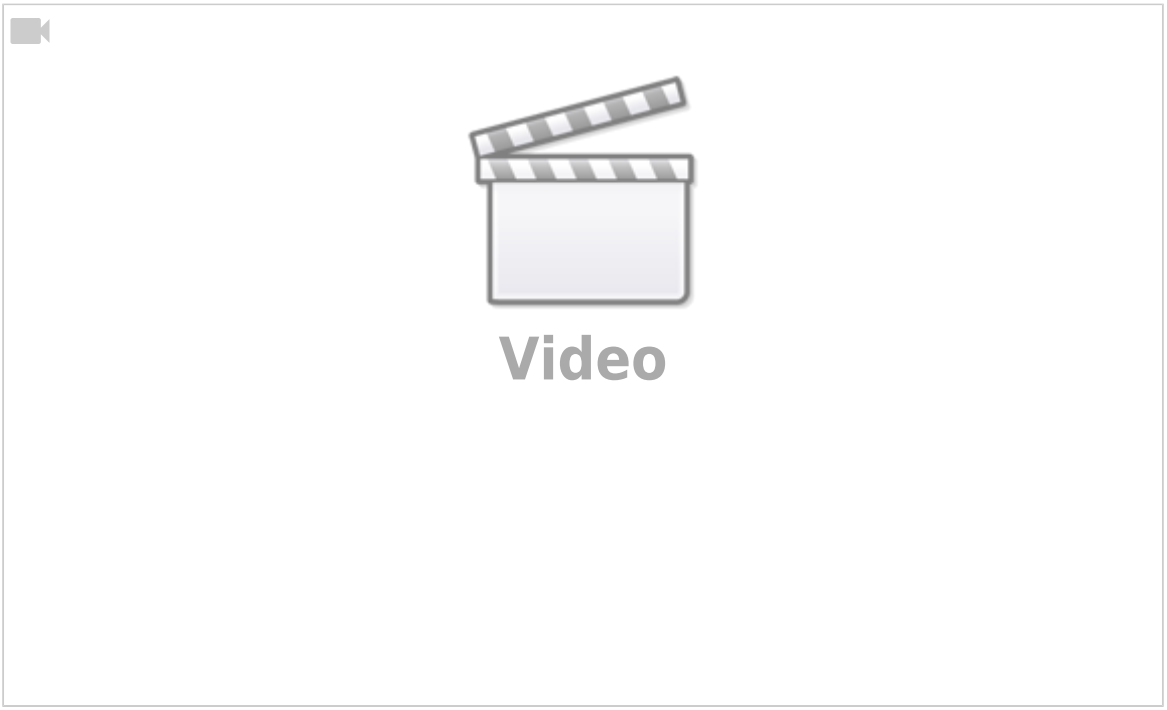
Most cloud services are accessible through a web browser, such as Google Chrome or Firefox, and some companies offer custom mobile applications. Some well-known cloud services examples include Google Drive, Netflix, Apple iCloud, Dropbox, and Microsoft.



How Cloud Works?

Information and data are stored on physical or virtual servers that a cloud computing service such as Amazon and it's AWS company retain and monitor. As a user of personal or business cloud computing, you use an internet connection to access the stored information on the cloud.

Here is a short video about how the cloud works in a simple manner;



Why Cloud Computing?

Cloud Computing evolved the ways we use a computer.

- From companies to private users, everybody relies on the cloud directly or indirectly most of the time in their daily lives.
- Nowadays, cloud-based activities are rising the Internet's capacity more than ever before. So, nearly everything in the digital world runs on cloud computing.
- It increases the value of the work and promises to reduce costs and helps users focus on their business and work rather than IT obstacles.
- It offers flexibility, data recovery, little or no maintenance, easy access and a higher level of security.

Advantages of The Cloud Technology

There are various benefits of cloud computing technology. The most important ones are given below.

Cost Efficiency

- One of the most important benefits of Cloud Computing is its economy.
- Cloud computing helps to reduce a significant amount of expenditure on both capital & operational manner.
- You do not need to invest in expensive hardware, storage devices, and software, and only pay for the services you use. This also saves the infrastructure costs and the money needed to manage the network.
- It provides the companies with the lowest possible level of operation with zero data capacity and software requirements, the business can save significant capital costs.

Elasticity and Flexibility

- Cloud computing helps you to reduce your resource demands and increase them according to your needs.
- For example, you can increase your resources if you have heavy traffic on your site and vice versa.
- Cloud computing gives you the flexibility to work anywhere you want, and all you need is an internet connection whenever you want.

Reliability

- Cloud computing is very reliable as the stored data is secured and can not be manipulated.
- Several copies of the data are being made, and if the database fails, the data from the other side can be recovered.
- The company can take advantage of both the vast pool of redundant IT services and the

process of failover.

Increased Security

- Everything you access and save with cloud computing is on the cloud. The providers of the service pick the highest level of data protection.
- Even if a laptop is lost or damaged, another computer can be used to access the company GUI. And since all of the records are stored on the cloud, there is no question about losing important documents because they have been saved on a hard drive laptop that is now lost or damaged.
- The full-time job of a cloud host is to track security carefully, which is significantly more efficient than a traditional in-house program, where an organization needs to divide its efforts among a multitude of IT issues, with security being just one of them.

Manageability

- Cloud computing provides improved and streamlined capabilities for IT management and maintenance by central resource management.
- Many items are handled by cloud computing. The only thing the user has to do is get an internet connection and a laptop.

Availability

- By its definition, cloud computing depends on the Internet, ensuring that businesses interested in starting or extending their use of cloud-based services need to work closely with an IT consulting firm to show them how to manage bandwidth rates that will be sufficient to meet their IT needs.
- Cloud service providers offer up to 99.99% uptime to ensure that business operations and executions continue to flow.

Centralization

- All data are stored in one location.
- So that multiple remote locations can be reached.

Auto-Updating

- Software updates and enhancements can be a painful thing that cloud computing simplifies for you.
- The cloud service provider looks after and controls all software maintenance and upgrades.

No Maintenance

- Organizations need to think about managing the entire system while operating a conventional server setup.
- A cloud computing solution eliminates the need for any maintenance.
- Not only does it increases work efficiency, but also reduces costs of operations in the longer run.

Disadvantages of The Cloud Technology

The drawbacks of cloud computing are as follows:

Internet Dependency

- Cloud computing requires internet connectivity as if you will not be able to access the cloud if there is no internet connection.
- There is no other way to access the data in the cloud.
- Similarly, a low-speed Internet connection makes cloud computing difficult and often impossible.

Downtime

- Cloud providers may face power loss, low internet connectivity, service maintenance, etc.
- A cloud outage is a period when cloud services are not available.
- So downtime or outage should also be considered while working with cloud computing.

Loss of Control

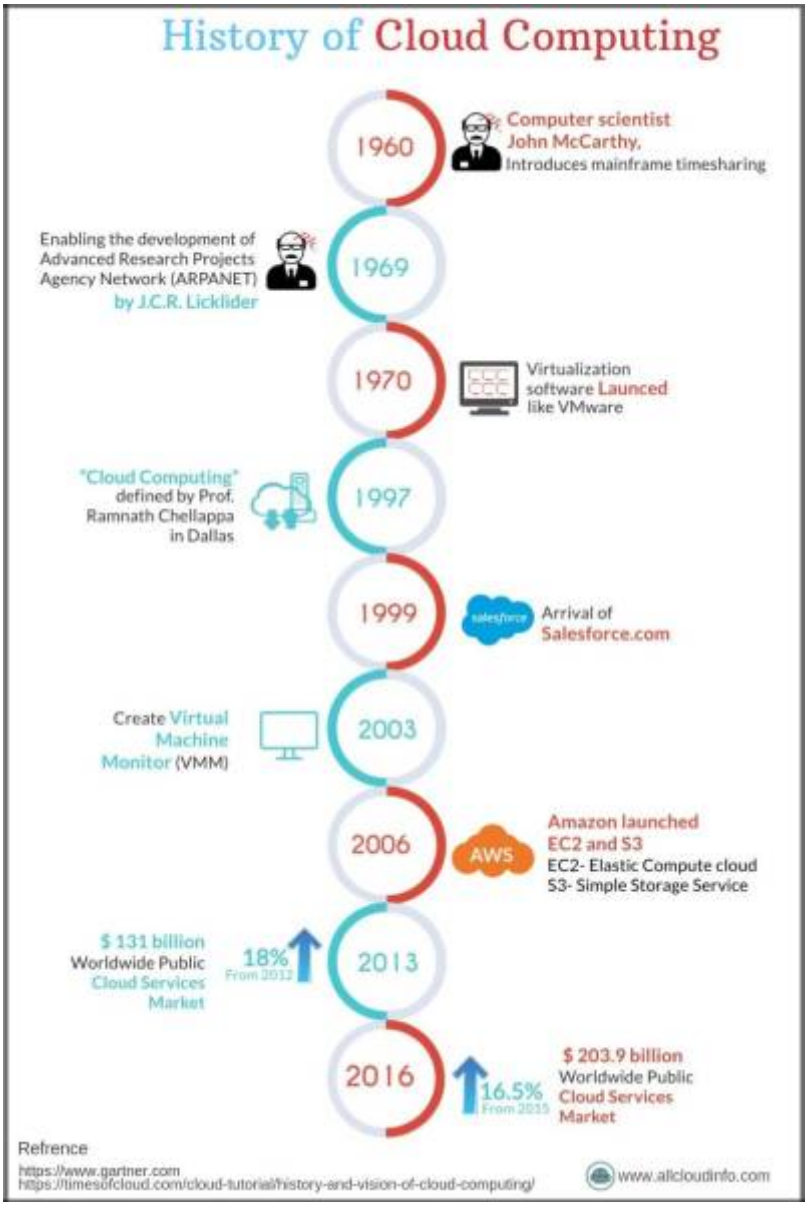
- Since cloud computing is very secure, it still requires attention.
- In essence, you trust another party to take care of your data.
- Once you accept cloud technology, you should be well aware that you will share all the sensitive information about your business with a third-party cloud computing service provider.

Lack of Support

- Cloud computing companies sometimes fail to provide customers with adequate support.
- If you have any technical problems, you have no choice but to call the technical support of your host provider for assistance.
- You can't fix the cloud computing problems, and some companies don't provide technical support around the clock.

Evolution of the Cloud Computing

When we think of cloud computing, we mostly look into the ideas and products we see in all-around and think that Cloud is all about the 21st century. But, in fact, Cloud concepts have existed long times ago, nearly the middle of the last century. Let's take a short tour around this period and try to understand the birth and evolution of Cloud computing.



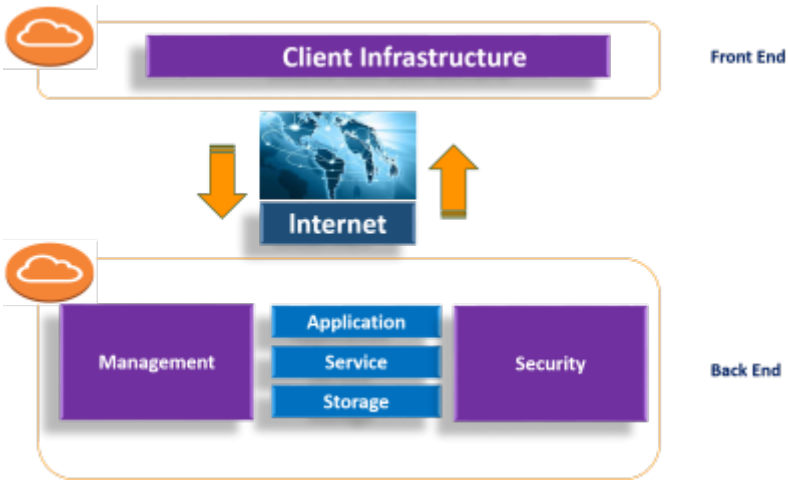
- The idea of cloud computing came into the picture in 1950 with accessible via thin/static customers and mainframe computer implementation.
- In 1959, Computer scientist John McCarthy initiates the first project to use a time-sharing system, which allows several people to use a single, central, computer at the same time.
- In 1969, J. C. R. Licklider, both a psychologist and a computer scientist, helped develop the ARPANET (Advanced Research Projects Agency Network), a “very” primitive version of the Internet. His vision was for everyone to be interconnected and accessing programs and data at any site like today's cloud computing.
- In 1970, the concept of virtualization has evolved with the Internet.
- In 1997, Professor Ramnath Chellappa from Emory University had mentioned the cloud in an article.
- In 1999, one of the first landmarks in the history of cloud computing was Salesforce.com's introduction of the idea of providing business applications through a single website.
- Amazon Web Services (AWS), which launched its public cloud in 2002, recognized the start of the modern-day cloud. At this point, there were virtually no competitors and while the advantages of using the cloud, such as elasticity and scalability, were recognized, the practical use cases were not yet available to persuade potential users.
- In 2006, Amazon launched Amazon Web Service (AWS) on a utility computing basis although the initial release dated back to July 2002. The most well-known of these services are Amazon EC2 and Amazon S3.

- In 2008, NASA's OpenNebula, enhanced in the RESERVOIR European Commission-funded project, became the first open-source software for deploying private and hybrid clouds, and for the federation of clouds.
- In April 2008, Google announced a preview release of App Engine, a developer tool that allowed users to run their web applications on Google infrastructure.
- In October 2008, Microsoft launched Azure.
- In 2011, IBM introduced the IBM SmartCloud framework, in support of Smarter Planet (a cultural thinking project). Then, Apple launched the ICloud, which focuses on storing more personal information (photos, music, videos, etc.).
- In 2012, Oracle launched the Oracle Cloud offering three business basics: IaaS (as-a-service infrastructure), PaaS (as-a-service platform) and SAAS (as-a-service software).

Cloud Computing Architecture

Parts of Cloud Computing Architecture

Cloud Computing architecture basically comprises of the two parts which are called Front-end and Back-end. Each of the ends is connected through a network, called Internet. The diagram below illustrates the architecture of the cloud computing:

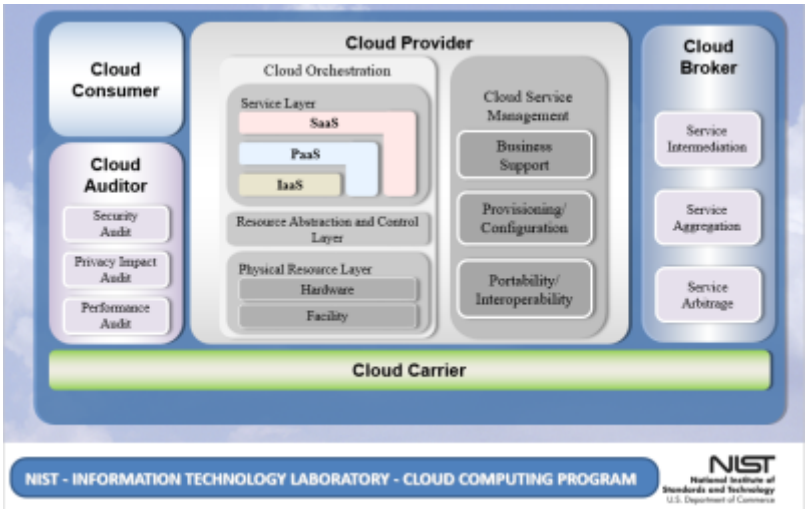


- The Front-end is the client part of Cloud Computing.
  - The front end is the end which is used by the user.
  - It includes the user interface and applications which are required to access the cloud computing platforms.
  - Example: Web Browsers.

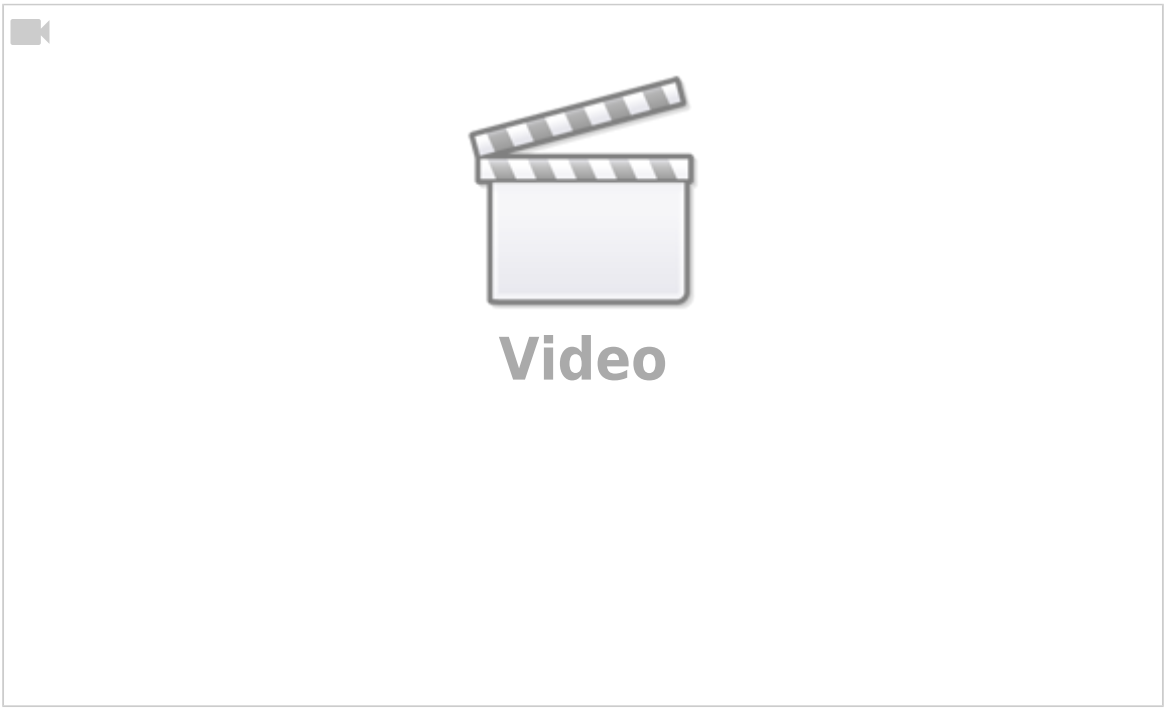
- The Back-end refers to the cloud itself.
  - The Back-end is managed by the host.
  - It consists of all resources which are necessary to provide cloud computing services such as virtual machines, data storage, deployment models, services models, security system, etc.
  - Providing built-in security mechanisms, traffic control and protocols is the responsibility of the Back-end.

Actor/Role Based Model

The following diagram by NIST (National Institute of Standards and Technology) shows the graphical view of cloud computing architecture actors in an actor/role-based model and the necessary architectural components for managing and providing cloud services such as service deployment, service orchestration, cloud service management, security, and privacy.

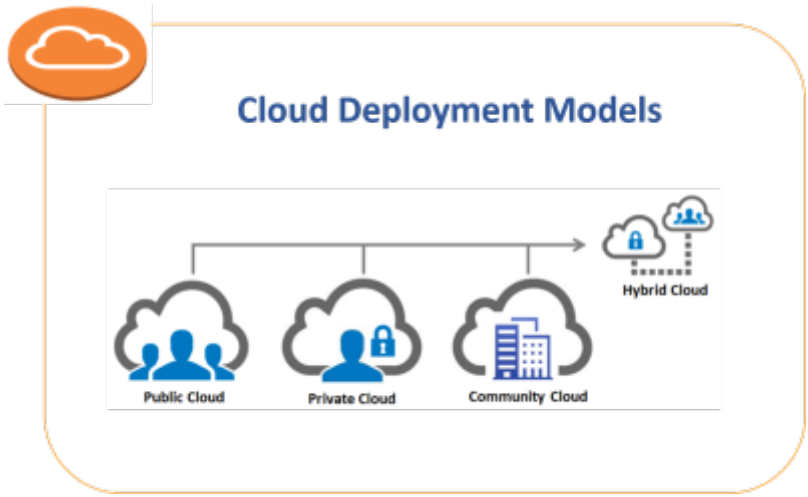


- A Cloud Consumer is an individual or organization that acquires and uses cloud products and services.
- The purveyor of products and services is the Cloud Provider.
- The Cloud Broker acts as the intermediate between consumer and provider and will help consumers through the complexity of cloud service offerings and may also create value-added cloud services as well.
- The Cloud Auditor provides a valuable inherent function for the government by conducting the independent performance and security monitoring of cloud services.
- The Cloud Carrier is the organization that has the responsibility of transferring the data akin to the power distributor for the electric grid.

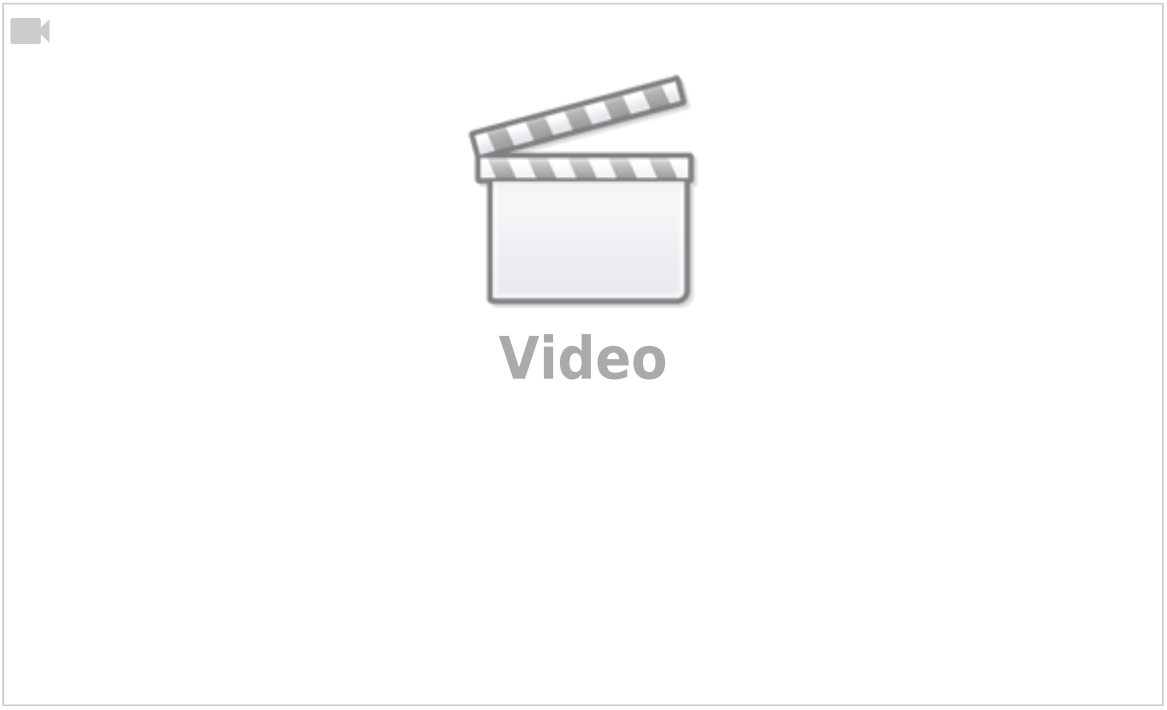


Deployment Models

Cloud Deployment Models



A cloud deployment model is a specific type of cloud environment, characterized primarily by ownership, scale, and access. Each deployment model is also defined by the location of the environment's infrastructure. There are four common cloud deployment models as seen in the image above.



Public Cloud

Public Cloud is the name of the information service used for platforms that transfer data to all individuals or organizations with internet access.

- It is a form of computing in which a service provider offers services through the internet to the public.

- Users do not need to purchase or update additional hardware or software.

Examples of Public Cloud Services:

- Amazon Elastic Compute Cloud (EC2)
- Google AppEngine
- Windows Azure Services Platform
- IBM Blue Cloud

Advantages of Public Cloud:

- Moving to an actual cloud infrastructure and using latest technologies
- Renting resources or applications at more reasonable prices
- Achieving high uptime through the use of reliable infrastructure, and having opportunities to work on infrastructure that has taken extra and unprecedented security precautions

Private Cloud

Private Cloud is a cloud computing service that provides the same Public Cloud benefits but uses private hardware dedicated to individuals, businesses or groups.

- It means using a cloud infrastructure (network) solely by one customer/organization.
- It is not shared with others, yet it is remotely located.
- In the cloud system, all data circulates, is inaccessible to public internet access.
- The system itself and limited access are configured by itself or by the service provider according to the conditions set by the user.
- Taking full advantage of the cloud infrastructure, Private Cloud provides greater control and security of resources.
- The security and control level is highest while using a private network.

An example of a private cloud deployment is where you maintain your own servers and infrastructure that hosts your applications and data. The key difference between private and public clouds is that you are not responsible for managing a public cloud hosting solution. A large company may select a private cloud, while a smaller company may select a public cloud.

Advantages of Private Cloud:

- It can be installed anyway and anywhere as you wish
- Prevents data loss with data storage system
- Reduces costs

Community Cloud

Community cloud means a shared platform, usually with shared data and data management considerations, between organizations.

- Although different companies do not choose this privacy and security cloud technology, it can be provided to people/organizations with more than one company to that of the intercompany data network.
- It is best suited to companies, business organizations, research organizations, and tenders.

- This helps group cloud users in first understanding and evaluating the market need.
- A community cloud, for example, may belong to a single country government and can be used by different departments of that government.
- It is possible to identify community clouds on and off the premises.

Like the private cloud solution, the community cloud has areas where it can greatly benefit agencies, but it’s also not the right solution in all circumstances.

Advantages of Community Cloud:

- Ability to easily share and collaborate
- Lower cost

Hybrid Cloud

Hybrid cloud means using both private and public clouds, depending on their purpose.

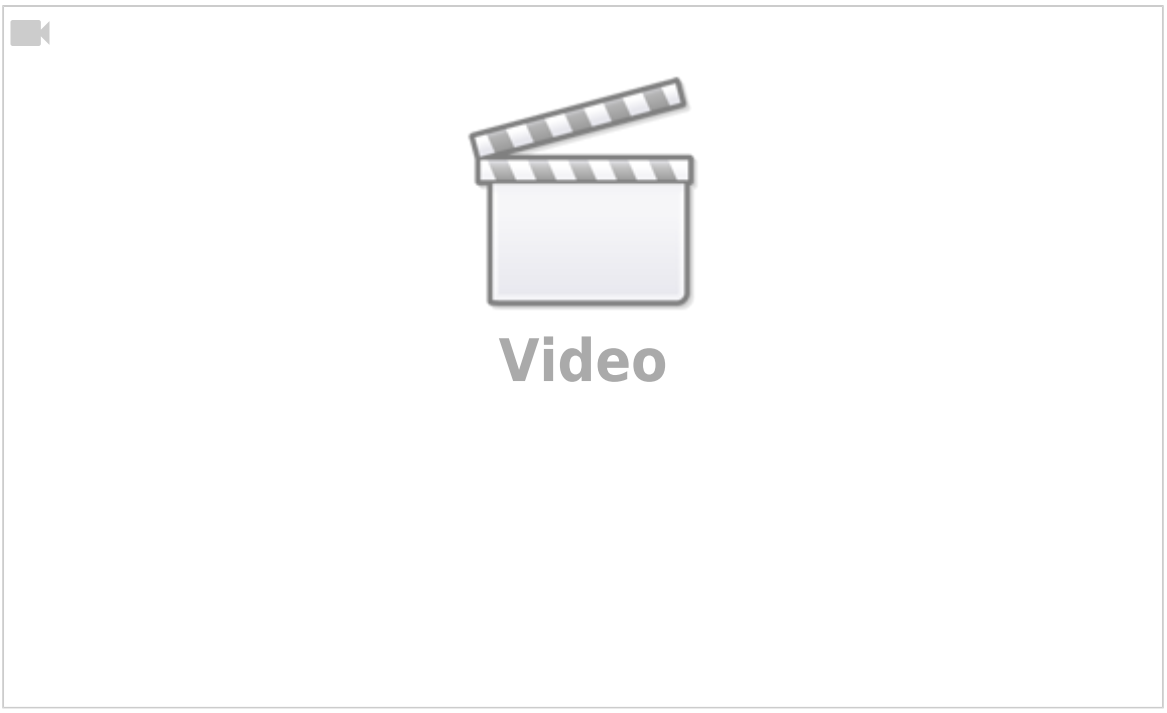
- It is a cloud technology used in situations where security and privacy are fundamental and require precautions.
- The non-critical activities are carried out using the public cloud, while the critical actions are carried out using a private cloud.
- For example, it can be used to interact with customers while retaining secure data via a private cloud.

Some hybrid clouds offer only a connection between the on-premise and public clouds. More and more businesses are moving to a managed hybrid cloud model that will mix and match dedicated infrastructure, private cloud systems, and public cloud resources to meet their unique needs.

Advantages of Hybrid Cloud:

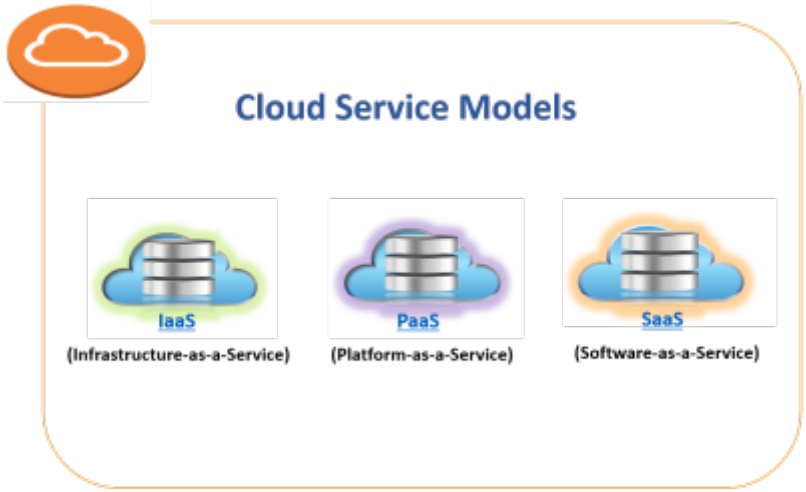
- The main advantage of a hybrid cloud model is control on the cloud .
- A hybrid cloud's scalability makes it an attractive alternative to a purely private cloud that can be incredibly expensive over time to upgrade and extend.
- Businesses can combine private cloud protection with public cloud resources and services.

Complementary Lesson about AWS Cloud Service Models



Service Models

Cloud Service Models



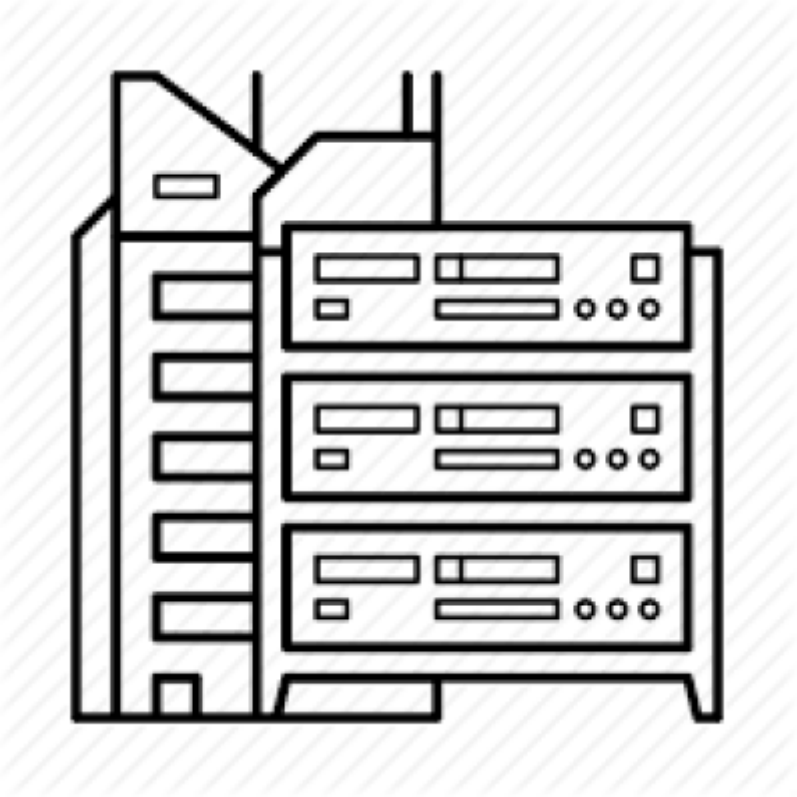
When you look more deeply into what resources a cloud architecture can provide, you start talking about models of cloud service. Cloud Computing technology basically offers a variety of service-based model. Cloud Service Models are all about how a cloud provider provides its services to customers. and the scope of customers' needs. There are three basic service models:

Now, let's examine these models one by one. But firstly, we will take a short look at the On-Premise term.

On-Premise

On-Premise is a server-based software. Because it is server based, it means that all archives and data

are displayed on the server source and not transferred to the internet environment.



- It is installed and runs on computers on the premises of the person or organization using the software, rather than at a remote facility such as a server farm or cloud.
- You are responsible for all the layers in the part where your software is installed on your or your company's computers.
- You must have a system team. This system team needs to know and set up servers, databases, security, network.
- It is necessary to make backups of the database and to install the current versions of the operating system.

IaaS - Infrastructure as a Service

Iaas, previously known as Hardware as a Service (HaaS), is a model based on a cloud computing platform. It's Cloud Computing's most basic service and the instant computing infrastructure which serves, manages, and monitors over the internet. For IaaS, a virtual server is built and users are equipped for cloud service.

- Customers outsource their IT infrastructure including servers, networking, distribution, storage, virtual machines, and other tools.
- The complete management is done by the Cloud Service provider.
- The installation, configuration, and management of the software are complete by the customer.
- Customers access these services on an Internet-based pay-per-use model, i.e. cloud computing network.
- It provides all computing resources but in a virtual environment so that multiple users can access them.
- The resources used can be increased/decreased at any time by taking advantage of the flexible structure of Cloud Computing.

IaaS supports companies of all shapes and sizes as it enables full control of your infrastructure and works on a pay-as-you-use model, making it ideal for most budgets.

PaaS - Platform as a Service

PaaS is a development framework for developers that is designed to create, test, run and manage applications for the programmer. A developer can easily write and deploy the application directly into this service.

- PaaS basically expands the IaaS layer by eliminating the virtual machine management problem.
- It not only includes server, storage, and networking but also database, tools, business services, and many more.
- PaaS helps developers of applications to build their projects by offering layers of hardware and software.
- This service includes system management, operating system, a framework for the programming language, database, etc.
- Because the service provider manages the system, you only manage applications and data.
- Some popular PAAS providers are AWS Elastic Beanstalk, Google App Engine, Microsoft Azure, etc.

When to Use PaaS PaaS is often the most cost-effective and time-effective way for a developer to create a unique application.

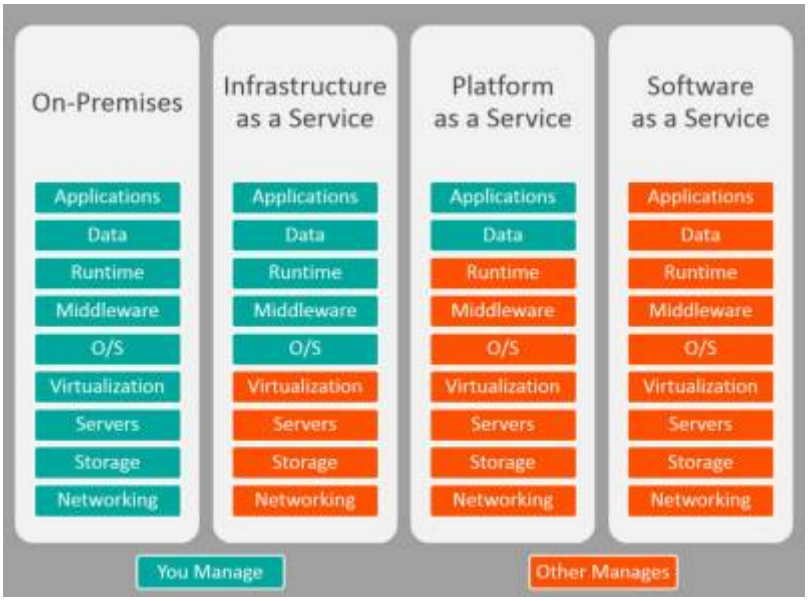
SaaS - Software as a Service

Software as a service (SaaS) is a software distribution model in which applications are managed by a third-party provider and made available to customers on the Internet.

- For many business applications such as business applications, it is one of the standard delivery models, including office software, messaging software, payroll processing software, and many more. Software as a Service (SaaS) is used by most leading organizations.
- Often known as host software, on-demand software, and web-based software are the SaaS applications.
- It allows users to connect to and use cloud-based applications over the Internet.
- E-mail, calendar, and office tools (such as Microsoft Office 365) are examples of these applications.

SaaS platforms are ideal for when you want an application to run smoothly and reliably with minimal input from you.

Comparison of Cloud Service Models



As you see, from On-premise to SaaS, the services offered to you are gradually increasing, and

ultimately almost everything you need is provided by the cloud provider.

- IaaS can be thought of as renting a computer through the cloud provider. It means that you have all the control here, including the operating system level, etc.
- In PaaS, the cloud provider here puts another level of abstraction for you to manage your needs easily. It expands the service frame and offers a wide range of solutions for applications. It saves you money and makes it easier for a distributed workforce to work together.
- In SaaS, software applications do not need to be installed and run on your device. Once you sign in to your account online, everything is available on the internet.

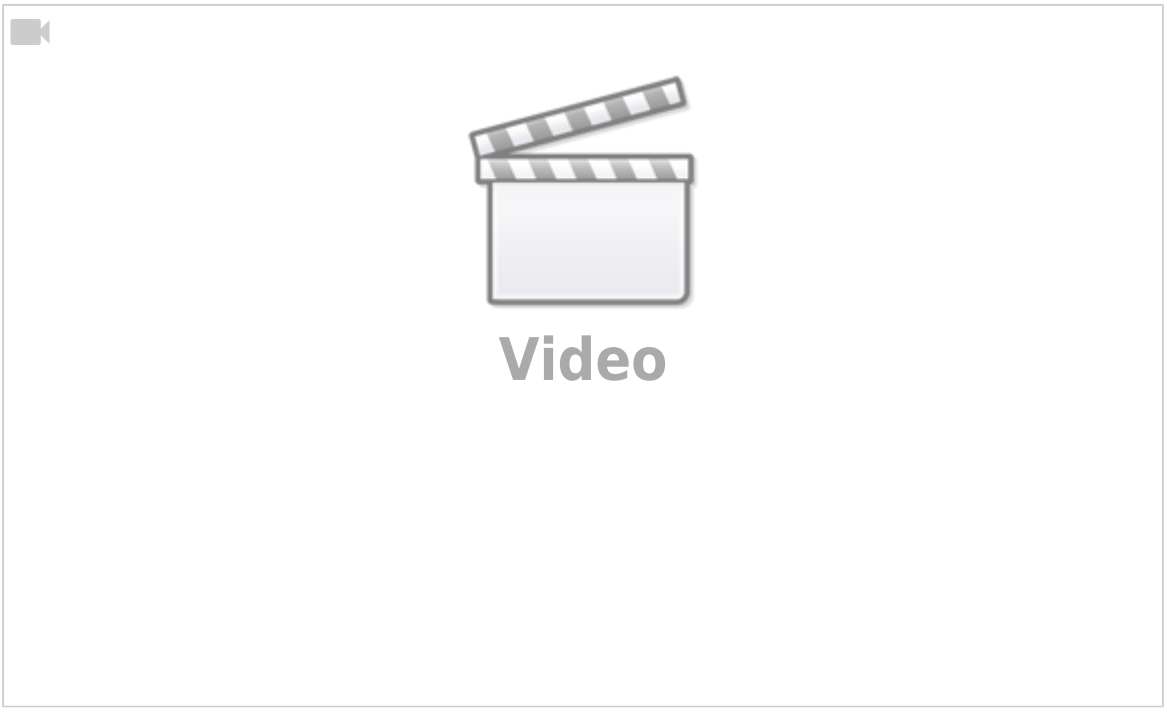
Virtualization

What is Virtualization?

Virtualization was first found by IBM in January 1967 as IBM Mainframe systems. The host computer is divided by specific software, serving multiple users simultaneously.

Virtualization refers to the operation of multiple operating systems called guests by sharing the same physical equipment resources.

- It allows you to use a physical machine's full capacity by distributing its capabilities among many users or environments.
- This will help the user to share a single physical resource instance or application with multiple users by providing multiple machines at the same time.
- The machine on which the virtual machine is to be created is called a host machine and the virtual machine is referred to as a guest machine.



Virtualization is software that renders computing environments independent of physical infrastructure, while cloud computing is a service that provides on-demand virtual computing resources (information and/or data).

- Cloud computing virtualization is a mechanism in which cloud users share the data in the cloud that can be device applications, etc.
- Virtualization systems create a logical layer between the user and the hardware, preventing the user from directly accessing physical system resources.
- This layer is a logical layer called Hypervisor or VMM (Virtual Machine Monitor), called the core of virtualization, that receives user requests and transmits them appropriately to the hardware.
- A hypervisor is synchronization between the server and the virtual environment and distributes resources through different virtual environments.

Main Virtualization Types

Since virtualization has so many varieties, it would not be right to limit the types of virtualization to a few areas. However, we can briefly define virtualization as 'doing more than one' and limit it to the most widely used types of business. Now we will focus on the main types of virtualization as below:

- Hardware Virtualization
- Software Virtualization
- Server Virtualization
- Storage Virtualization
- Operating System - OS Virtualization

Hardware Virtualization

Hardware virtualization is a type of virtualization that enables computers to be virtualized and moved to the data center and then to be used by end-users by accessing these resources over the network or the internet. It is the virtualization of computers as full hardware platforms, other logical abstractions of their part, or only the features needed to run different operating systems.

The hardware on the existing physical system is produced more than one virtual and provides complete isolation between guest operating systems. While this is the case, you can add, install and run guest (virtual) operating systems running on real hardware. The hypervisor layer will produce virtual hardware and communicate with each other.

Software Virtualization

Virtualization is done by special software installed on the operating system. The installed software creates one special file and keeps it for use by operating systems that can be run by the software.

The software runs the selected operating system exactly as if it were installing an operating system on a computer and shows it to you in the graphical interface.

Without the need to install software on a client, it is the fulfillment of the requirements of software on the server.

Virtualized software is an application that is loaded in its own network. Example of software virtualization is VMware, virtual box, etc.

Server Virtualization

Virtualizing the server is partitioning a physical server into multiple virtual servers and used to optimize the efficiency of the server. Physical server partitioning software is used in many virtual environments, called virtual servers or private servers.

Each virtual server runs its own program and operating system. We can have a number of virtual servers on the same computer instead of having a separate computer for each web server.

Storage Virtualization

Physical storage is partitioned by logical storage and abstracted between each other. It refers to the retention of data in a virtualized file.

It can also be named as a category of an available storage unit that manages simply from a central console. Device operating systems and applications can use the disks for writing directly on their own.

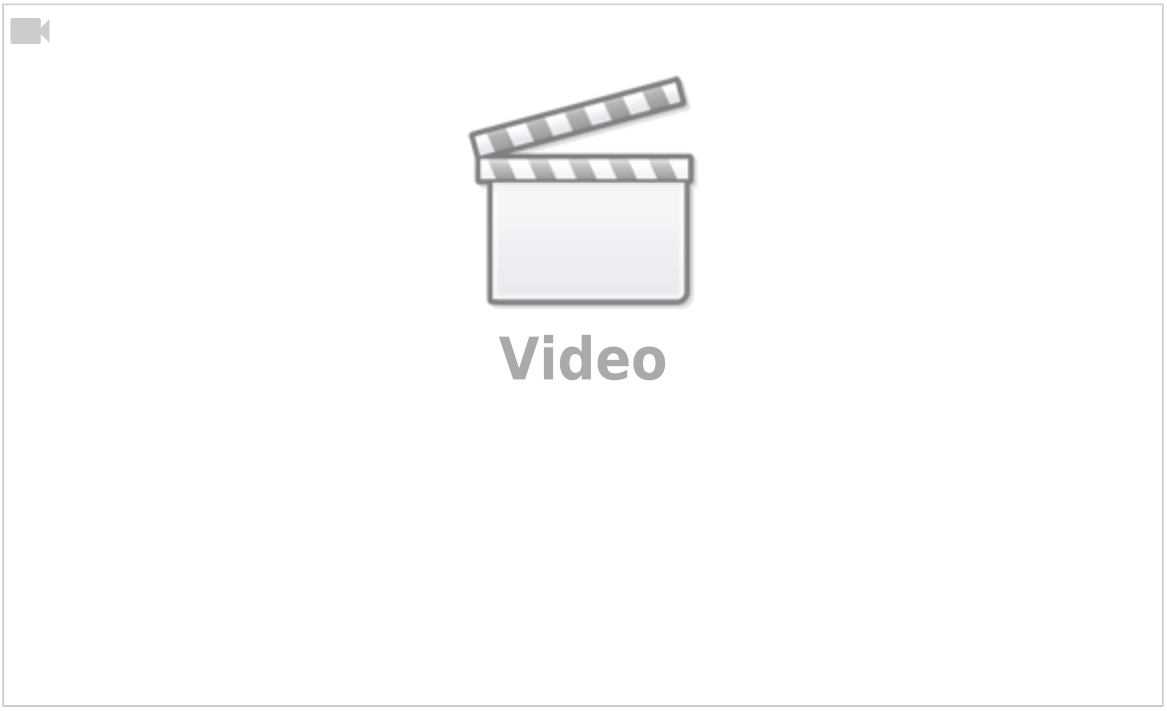
Such virtualization provides numerous advantages such as fast data backup, accomplishment, and recovery.

Operating System - OS Virtualization

It shares the operating system in common data sources, providing the advantage of maintenance, repair, and enhanced system security in other users' operating systems. It provides isolation of those who use virtualization software in the same way.

Nothing is pre-installed or permanently loaded on the local device with the help of OS virtualization and no hard disk is needed. Using a kind of virtual disk, all run from the network. The client connects to this virtual disk via the network and boots with the virtual disk enabled operating system.

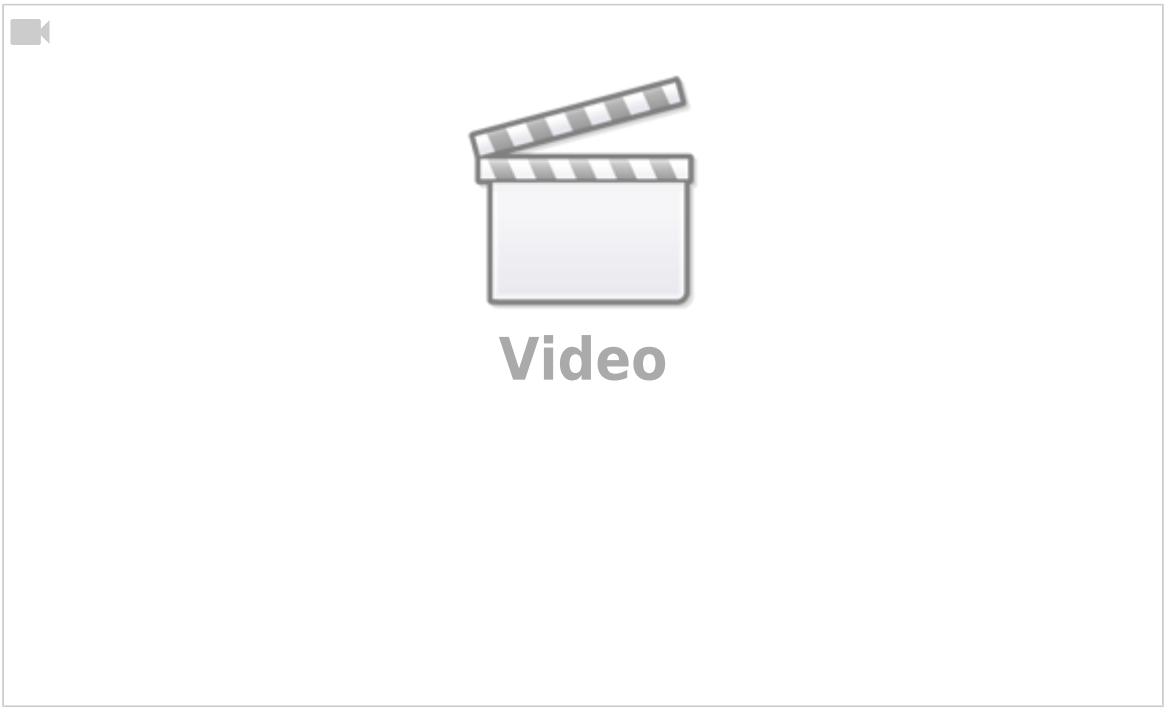
Complementary Lesson about Hypervisor



Amazon Web Services

Introduction to AWS

What is AWS?



AWS stands for Amazon Web Services that offers various IT services on demand using distributed IT infrastructure and offers flexible, reliable, scalable, and cost-effective cloud computing solutions.

Amazon Web Services (AWS) is the world’s most comprehensive and broadly adopted cloud platform, offering nearly 200 fully-featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—are using AWS to lower costs, become more agile, and innovate faster.

AWS provides different services such as infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and a wide range of workloads, including game development, data processing, warehousing, performance growth, and more.

History of AWS

AWS was invented by Amazon.com and it's the origin as a developer platform can be traced back to 2002 when an initial beta called Amazon.com Web Service was launched.

In 2003, when Chris Pinkham and Benjamin Black presented a paper describing a vision of Amazon's retail computing infrastructure, the AWS concept was publicly reformulated.

In 2004, the first AWS service launched for public usage: Simple Queue Service (SQS).

In 2006, Amazon Web Services LLC (AWS), an Amazon.com company, officially began offering developer customers access to in-the-cloud infrastructure services based on Amazon's own back-end technology platform. The three initial service offerings of Amazon S3 cloud storage, SQS, and EC2 were combined.

In 2007, Amazon launched Amazon SimpleDB, which enables organizations, academics, data analysts and developers to process large amounts of data easily and cheaply.

In 2008, Amazon announced Elastic IPs, Amazon Elastic Block Store (EBS), Amazon CloudFront,

In 2009, Amazon launched Amazon Elastic MapReduce (EMR), Elastic Load Balancing (ELB), Virtual Private Cloud (VPC), Amazon Relational Database Service (RDS), EC2 Spot Instances, Amazon Route 53, AWS Elastic Beanstalk,

In 2010, it was announced that all the retail sites of Amazon.com had moved to AWS. AWS launched Simple Notification Service and AWS CloudFormation,

In 2011, AWS announced the launch of Amazon Simple Email Service (SES),

In 2012, AWS launched Amazon DynamoDB, AWS Identity and Access Management (IAM) for EC2, Amazon Glacier, Amazon Redshift,

In 2013, AWS announced AWS CloudTrail, released Amazon Kinesis, AWS Lambda,

In 2014, AWS announced Amazon Aurora, EC2 Container Service (ECS),

In 2015, AWS launched AWS API Gateway Service, AWS Elasticsearch Service, Snowball, Internet of Things platform, Amazon Elastic Container Registry (ECR).

In 2016, AWS announced Auto Scaling for Amazon EC2 Container Service (ECS), Elastic File System

(EFS), AWS Snowmobile, Snowball Edge, Amazon Lightsail and AWS acquired Cloud9.

In 2017, AWS announced Amazon Glue, Amazon SageMaker, AWS CloudWatch agent.

In 2018, AWS announced AWS Elastic Kubernetes Service (EKS).

AWS Features

Amazon Web Services has a variety of features that make it consistent across different companies. AWS's characteristics are:

Easy to use:

AWS is designed to allow application providers, ISVs, and vendors to quickly and securely host your applications – whether an existing application or a new SaaS-based application. You can use the AWS Management Console or well-documented web services APIs to access AWS's application hosting platform.

Reliable:

With AWS, you take advantage of a scalable, reliable, and secure global computing infrastructure, the virtual backbone of Amazon.com's multi-billion dollar online business that has been honed for over a decade.

Flexible:

AWS enables you to select the operating system, programming language, web application platform, database, and other services you need. With AWS, you receive a virtual environment that lets you load the software and services your application requires. This eases the migration process for existing applications while preserving options for building new solutions.

Scalable and high-performance:

Using AWS tools, Auto Scaling, and Elastic Load Balancing, your application can scale up or down based on demand. Backed by Amazon's massive infrastructure, you have access to compute and storage resources when you need them.

Cost-Effective:

You pay only for the compute power, storage, and other resources you use, with no long-term contracts or up-front commitments. For more information on comparing the costs of other hosting alternatives with AWS, see the AWS Economics Center.

Secure:

AWS utilizes an end-to-end approach to secure and harden the infrastructure, including physical, operational, and software measures. For more information, see the AWS Security Center.

Complementary Lesson about AWS Services Group

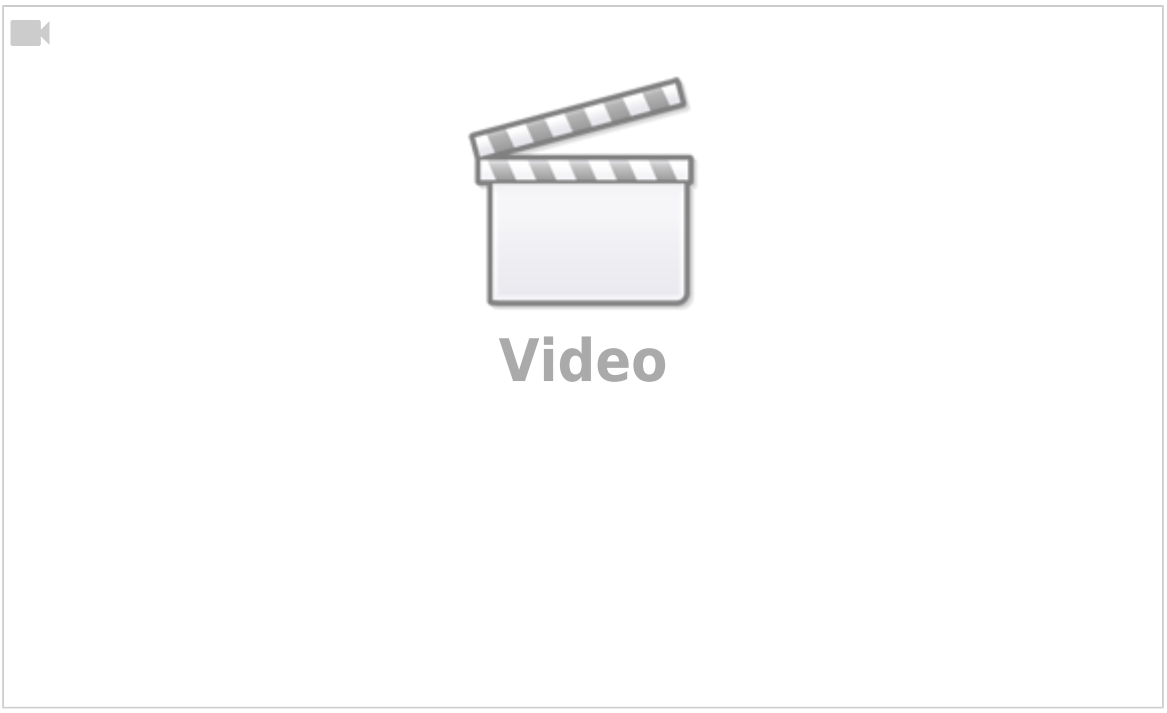
https://www.youtube.com/watch?v=mZ5H8sn_2ZI&t=3s

AWS Infrastructure

Introduction

AWS serves over a million active customers in more than 190 countries. The AWS Cloud infrastructure is built around AWS Regions and Availability Zones.

- An AWS Region is a physical location in the world where it has multiple Availability Zones.
- Availability Zones consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities.
- These Availability Zones offer you the ability to operate production applications and databases that are more highly available, fault-tolerant, and scalable than would be possible from a single data center.
- The AWS Cloud operates in over 80 Availability Zones within over 25 geographic Regions around the world, with announced plans for more Availability Zones and Regions (as of March 2021).
- For more information on the AWS Cloud Availability Zones and AWS Regions, see AWS Global Cloud Infrastructure and AWS Global Infrastructure.



Amazon Regions

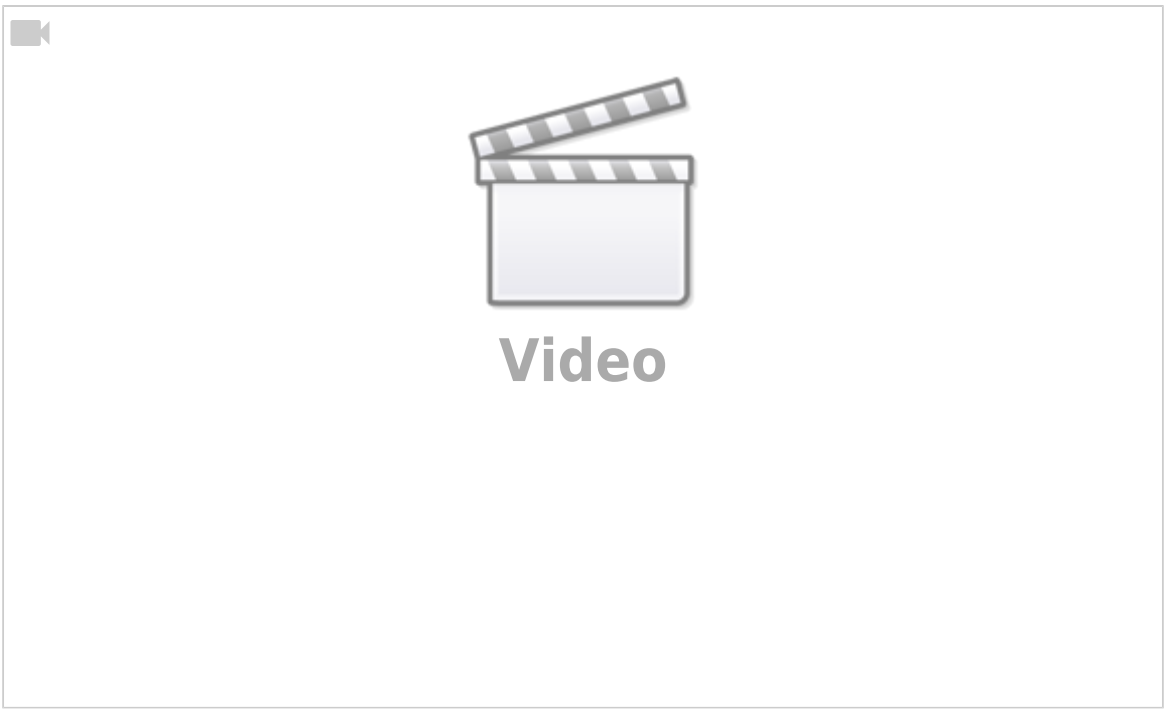
AWS has the concept of a Region, which is a physical location around the world where it clusters data centers.

- Each group of logical data centers is called Availability Zone (AZ).
- Each AWS Region consists of multiple, isolated, and physically separate AZ's within a geographic area.
- Unlike other cloud providers, who often define a region as a single data center, the multiple AZ design of every AWS Region offers advantages for customers.
- Each AZ has independent power, cooling, and physical security and is connected via redundant, ultra-low-latency networks. AWS customers focused on high availability can design their applications to run in multiple AZ's to achieve even greater fault-tolerance.
- AWS infrastructure Regions meet the highest levels of security, compliance, and data protection.

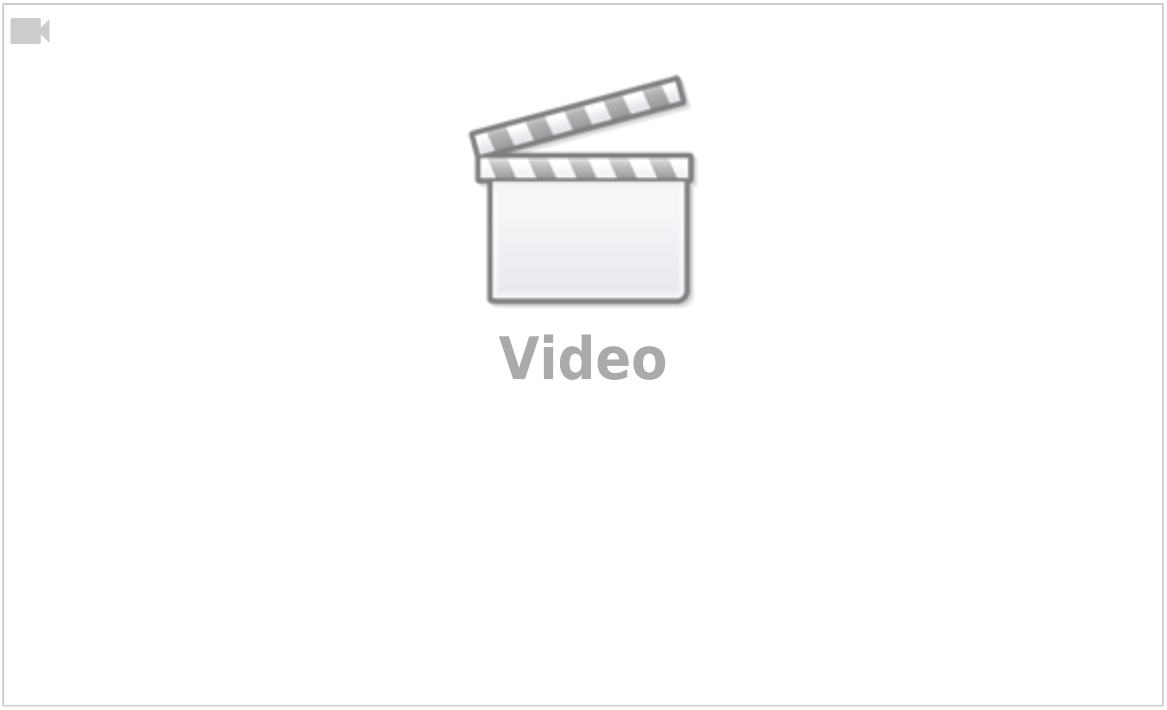
Video

AWS Availability Zones

- An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region.
- AZ's give customers the ability to operate production applications and databases that are more highly available, fault-tolerant, and scalable than would be possible from a single data center.
- All AZ's in an AWS Region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fiber providing high-throughput, low-latency networking between AZ's.
- All traffic between AZ's is encrypted.
- The network performance is sufficient to accomplish synchronous replication between AZ's.
- AZ's make partitioning applications for high availability easy. If an application is partitioned across AZ's, companies are better isolated and protected from issues such as power outages, lightning strikes, tornadoes, earthquakes, and more.
- AZ's are physically separated by a meaningful distance, many kilometers, from any other AZ, although all are within 100 km (60 miles of each other).



AWS Data Centers



A data center is a facility that centralizes the IT processes and resources of an enterprise, as well as where the data are processed, handled and disseminated.

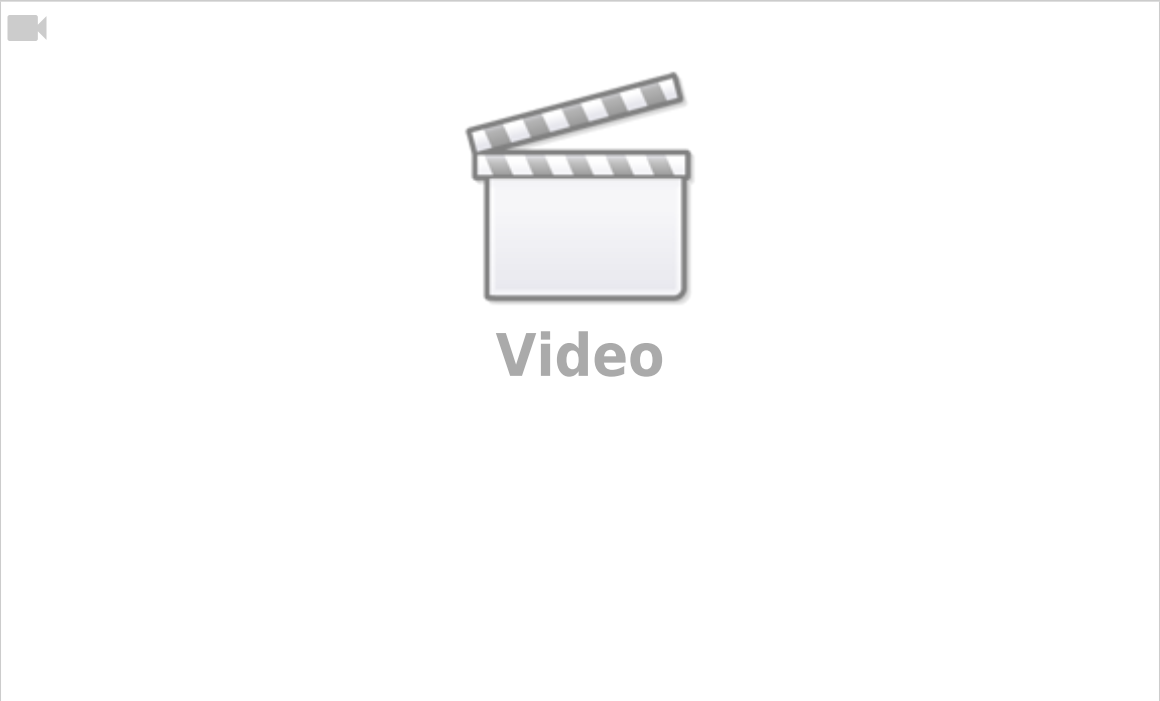
AWS pioneered cloud computing in 2006, creating cloud infrastructure that allows you to securely build and innovate faster. AWS is continuously innovating the design and systems of data centers to protect them from man-made and natural risks.

Edge Locations

An edge location is where end-users access services located at AWS and used for caching content.

Edge locations serve requests for CloudFront and Route 53.

- CloudFront is a content delivery network, while Route 53 is a DNS service.
- Requests going to either one of these services will be routed to the nearest edge location automatically. This allows for low latency no matter where the end-user is located.
- They are located in most of the major cities around the world and are specifically used by CloudFront (CDN) to distribute content to end-user to reduce latency.
- It is like a frontend for the service we access which are located in the AWS cloud.



AWS Local Zones

AWS Local Zones place compute, storage, database, and other select AWS services closer to end-users.

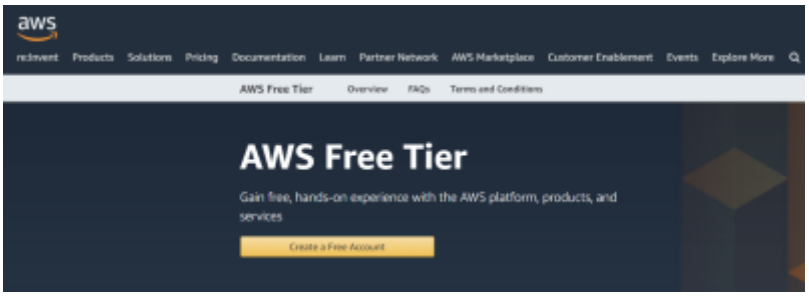
- With AWS Local Zones, you can easily run highly-demanding applications that require single-digit millisecond latencies to your end-users such as media & entertainment content creation, real-time gaming, reservoir simulations, electronic design automation, and machine learning.
- Each AWS Local Zone location is an extension of an AWS Region where you can run your latency-sensitive applications using AWS services such as Amazon Elastic Compute Cloud, Amazon Virtual Private Cloud, Amazon Elastic Block Store, Amazon File Storage, and Amazon Elastic Load Balancing in geographic proximity to end-users.
- AWS Local Zones provide a high-bandwidth, secure connection between local workloads and those running in the AWS Region, allowing you to seamlessly connect to the full range of in-

region services through the same APIs and toolsets.

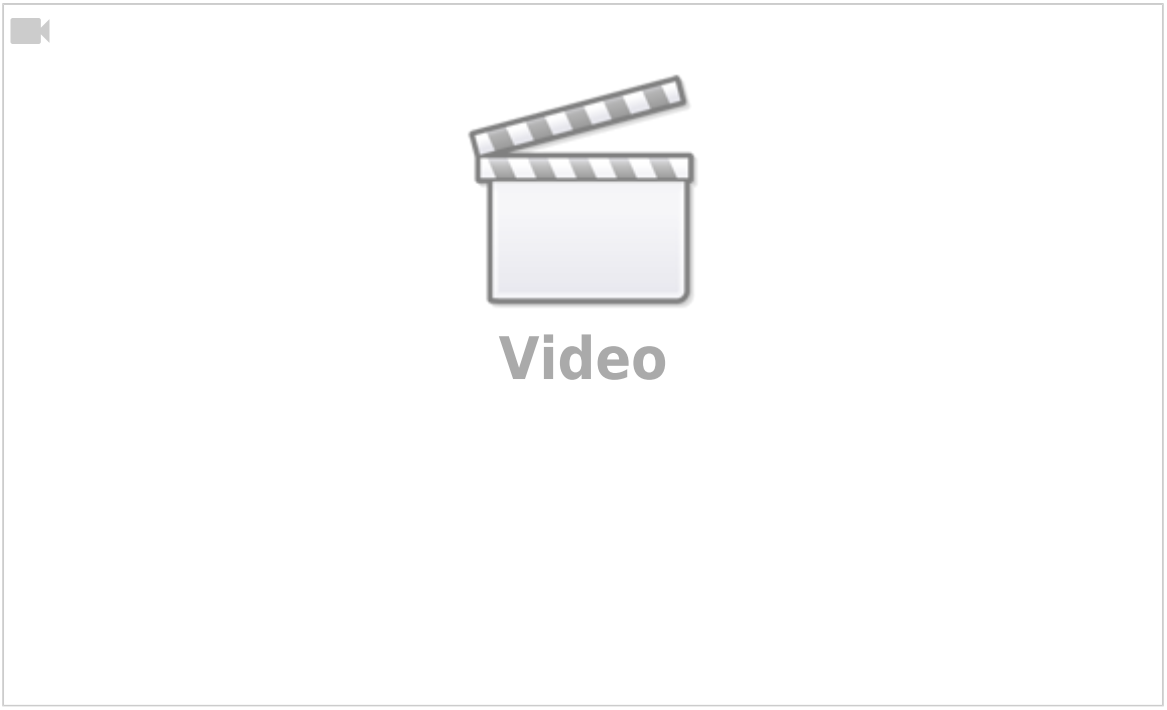
AWS Free Tier

What is AWS Free Tier?

The AWS Free Tier provides customers the ability to explore and try out AWS services free of charge up to specified limits for each service. You can explore more than 85 products and start building on AWS using the free tier.



- The Free Tier is designed to give you hands-on experience with a range of AWS services at no charge.
- For example, you can explore AWS as a platform for your business by setting up a test website with a server, alarms, and database.
- You can also try out services for developers, such as AWS CodePipeline, AWS Data Pipeline, and AWS Device Farm.
- If you don't use the full benefits provided by the Free Tier in a given month, the benefits don't roll over to the next month.
- To maximize your Free Tier benefits, be sure to spend time with AWS each month, trying out the services that you're curious about.
- For more detailed information, please follow this link.



Types of Offers

The Free Tier is comprised of three different types of offerings, a 12-month Free Tier, an Always Free offer, and short term trials. AWS Free Tier

- Services with a 12-month Free Tier allow customers to use the product for free up to specified limits for one year from the date the account was created.
- Services with an Always Free offer allow customers to use the product for free up to specified limits as long as they are an AWS customer.
- Services with a short term trial are free to use for a specified period of time or up to a one-time limit depending on the service selected.

The 12 Month Free Tier is only available to new AWS customers and is available for 12 months following your AWS sign-up date. The Other Offers are available to both existing and new AWS customers and may be limited in duration (such as for trials) or in available free usage (such as the amount of free storage for a database Offer).

Billing Policy

To avoid charges while on the Free Tier, you must keep your usage below the Free Tier limits.

- You are charged for any usage that exceeds the limits.
- To help you stay within the limits, you can track your Free Tier usage and set a billing alarm to notify you if you start incurring charges.
- If your application use exceeds the free tier limits, you simply pay standard, pay-as-you-go service rates.
- AWS Free Tier is applied to your monthly usage. It will expire on the 1st day of each month and does not accumulate.

You can see current and past usage activity by service and region by logging into your account and going to the Billing & Cost Management Dashboard.

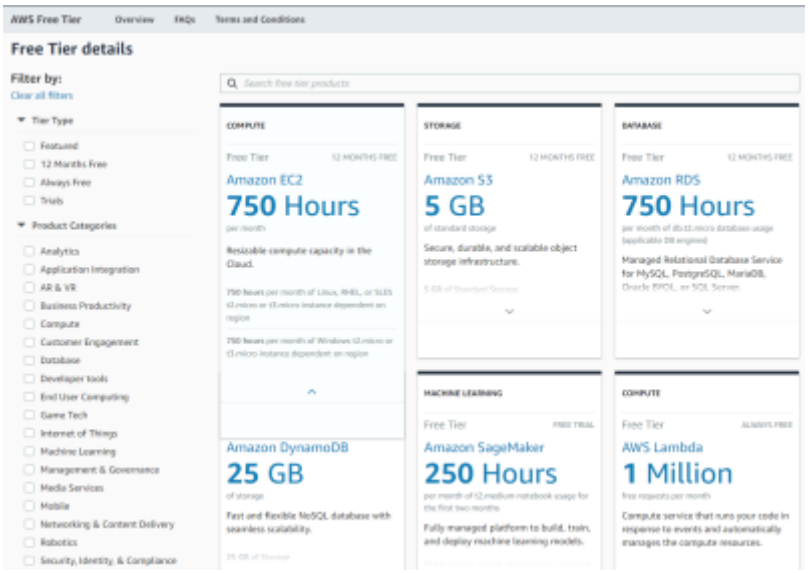
Limits

All services that offer a Free Tier have limits on what you can use without being charged. Many services have multiple types of limits.

- For example, Amazon EC2 has limits on both the type of instance you can use and how many hours you can use in one month.
- Amazon S3 has a limit on how much storage you can use and on how often you can call certain operations each month.
- For example, the Free Tier covers the first 20,000 times you retrieve a file from Amazon S3, but

you're charged for additional file retrievals.
- Each service has limits that are unique to that service.

Some of the most common limits are by time, such as hourly or by the minute, or by requests, which are the requests you send to the service, also known as API operations.

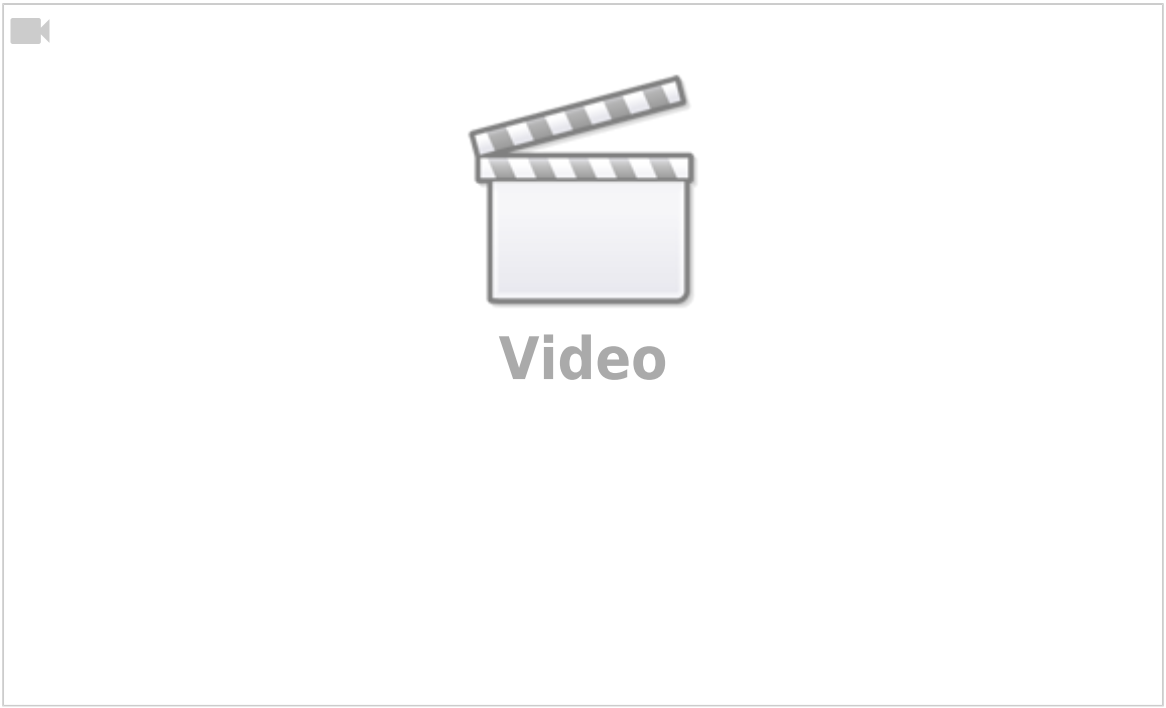


For more information about Free Tier limits, see AWS Free Tier.

Creating AWS Account

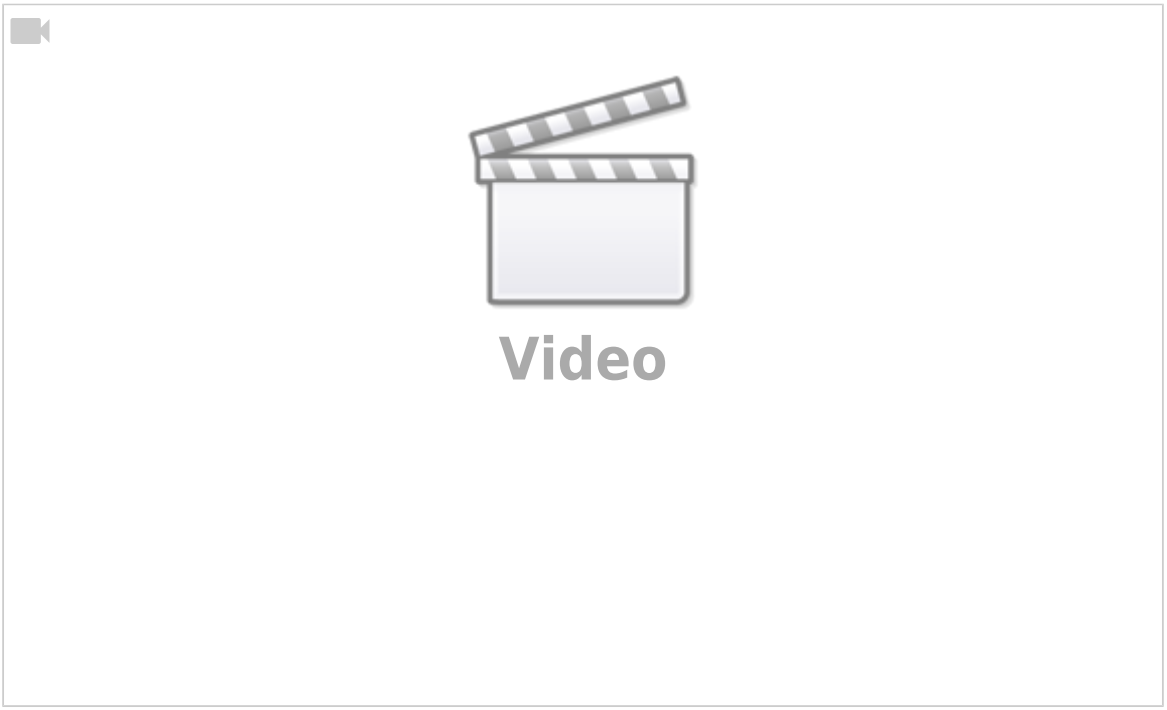
Creating a Free Tier Account

Here, we will learn how to create and activate the AWS Account.

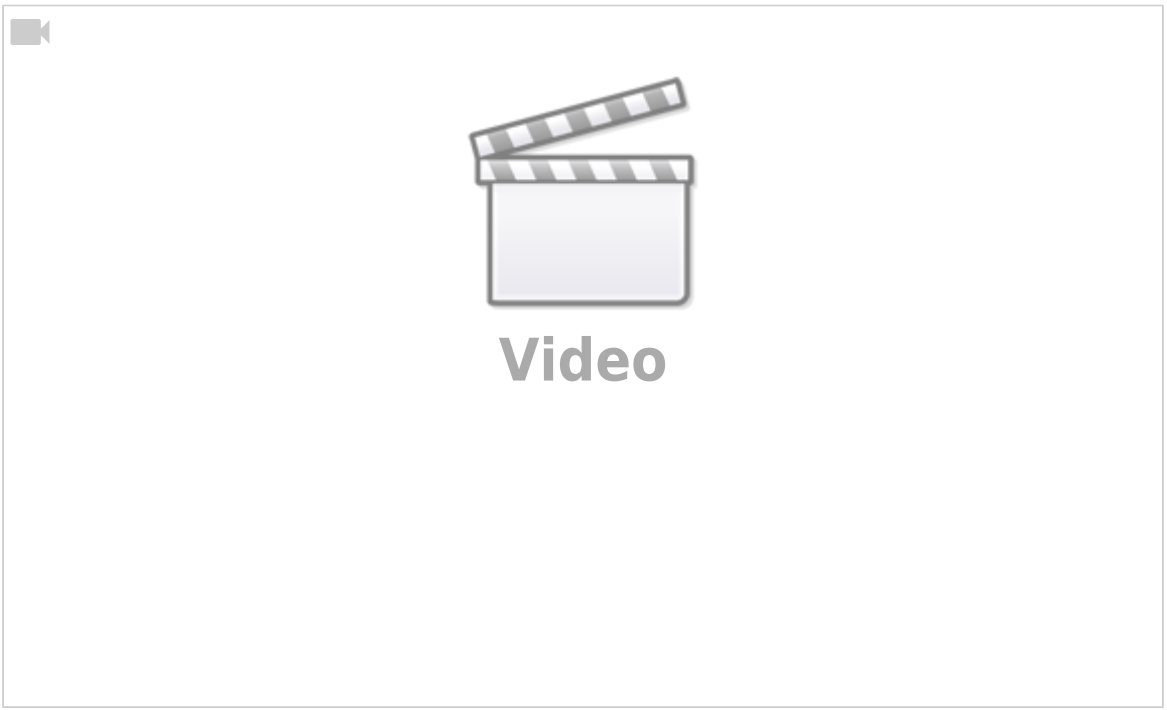


Setting an AWS Billing Alarm

Although we use a free account, there are some limitations to the free policy due to the free tier features. Therefore, setting up a billing alarm will allow us to be aware of any unintended charges.



Checking Free Tier Usage



Amazon IAM - Identity & Access Management

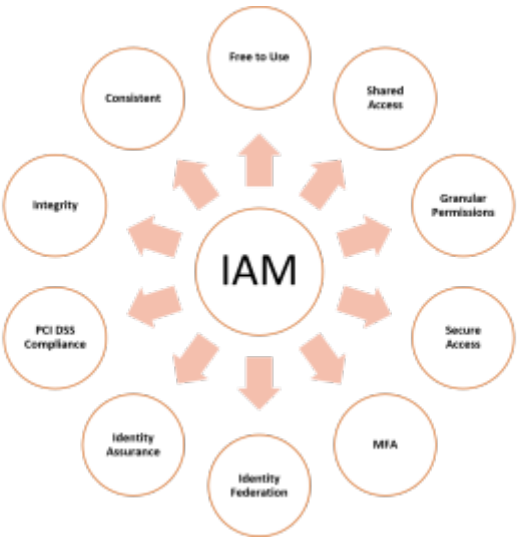
Introduction to IAM

What is IAM?

AWS IAM stands for Identity & Access Management and is the primary service that handles authentication and authorization processes within AWS environments. As an AWS account management service, it lets you control access to AWS services in a secure manner and helps to monitor who is authenticated and allowed to use resources.

- By using AWS IAM, you can manage users and their access level.
- All account settings are made through this service.
- It allows us to create and manage objects such as User, Group, Role, and Policy.
- Account owner can identify and allow the user to use specified services.
- All kinds of user password restrictions and multifactor authentication settings are also made through IAM.

IAM Features



- Free to use:

AWS IAM is an AWS account feature which is offered at no extra charge. By using IAM, you will only be paid when you use other AWS services.

- Shared access to your AWS account:

Users can share the resources for collaborative projects among themselves. You can also allow other users in your AWS account to manage and use services without having to share your password or access key.

- Granular permissions:

Different people can be granted permissions for different resources. So, it also means setting the authorization for the user to use a specific service but not others.

- Secure access to AWS resources for applications that run on Amazon EC2:

The IAM features can be used to provide secure credentials to applications running on EC2 instances.

- Multi-factor authentication (MFA):

AWS offers multifactor authentication to sign in to the AWS Management Console. Users not only need to provide a password or access key to work with your account, but also a code from a device that has been configured specifically.

- Identity federation:

You can allow users who already have passwords elsewhere - such as Twitter, Facebook, Linkedin - to access your AWS account temporarily. Users can log in to the AWS Console with the same username and password as they log in to Facebook, Twitter, etc.

- Identity information for assurance:

You receive log records that contain information about those who have made resource requests based on IAM identities.

- PCI DSS Compliance:

IAM supports the processing, storage, and transmission of credit card data by a merchant or service provider, and has been validated as being compliant with Payment Card Industry (PCI) Data Security Standard (DSS).

- Integrated with many AWS services:

IAM is integrated with many different AWS services.

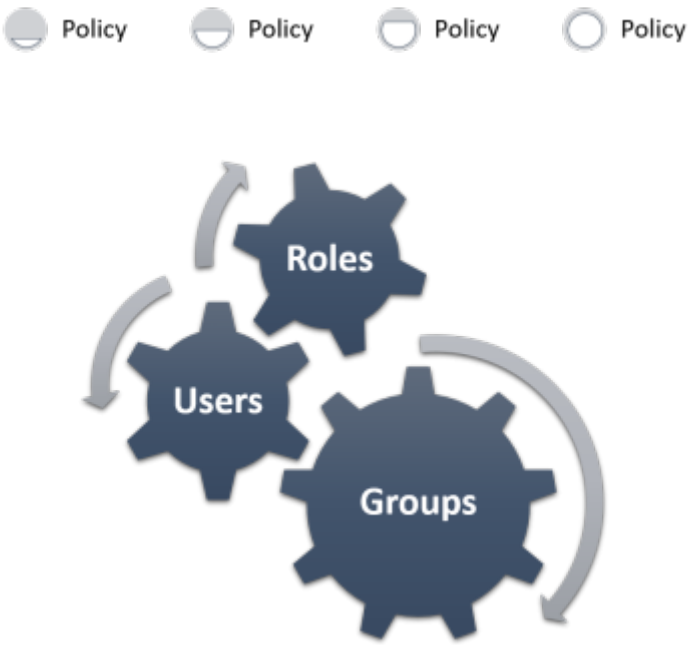
- Eventually Consistent:

Basically, IAM service is fairly consistent because it achieves high availability by replicating the data through multiple servers in the Amazon data center around the world.

For a more detailed explanation, you can follow this link.

Categorizing IAM Components

IAM components can be mainly categorized under two terms; identities and permissions.



IAM identities are created to give people and processes authentication in an AWS account. There are three identities in AWS IAM:

- Users
- Groups
- Roles

Permissions can be defined as different types of policies that use authorization to users.

- Policies

Security Credentials

While interacting with the AWS as any kind of user, you can have different kinds of security credentials that depend on how you communicate with the AWS. For example; you can use an email and password when logging in to the AWS console as a root user, whereas when logging in as an IAM user, you use a username and password. These combinations are called Security credentials in AWS.

Here are the security credentials you will commonly use when interacting with the AWS resources :

E-mail address and password is used when connecting to the AWS Management Console with the root account.

IAM username and password is used for accessing the AWS Management Console as an IAM user.

Multifactor authentication (MFA) is an additional layer of security that can be used with both the root account and IAM user as well.

Access keys are used with the CLI, APIs, and SDKs.

IAM - Users

What is an IAM user?

An IAM user is an entity that you create in AWS.

- The IAM user represents the person or service who uses AWS services.
- A primary use for IAM users is to give people the ability to sign in to the AWS Management Console for interactive tasks and to make programmatic requests to AWS services using the API or CLI.
- A user in AWS consists of a name, a password to sign in to the AWS Management Console, and up to two access keys that can be used with the API or CLI.
- When you create an IAM user, you grant it permissions by making it a member of a group that has appropriate permission policies attached (recommended), or by directly attaching policies to the user.
- You can also clone the permissions of an existing IAM user, which automatically makes the new user a member of the same groups and attaches all the same policies.

IAM User Types

An IAM user is an identity that has an associated credential and the permissions attached. This could be a real person who is a user or a web application, service account, auditing or back-up software. Firstly, we will focus on users who are real persons.

We created an AWS Free Tier account at the end of the previous section. This account is defined as a root user in the AWS world. Now let's continue to get to know the features of the IAM service, starting with the root user.

Account Root User

By first creating an AWS account, you create a root user identity account that is used to log in to the AWS. This identity is called the AWS Account Root User.

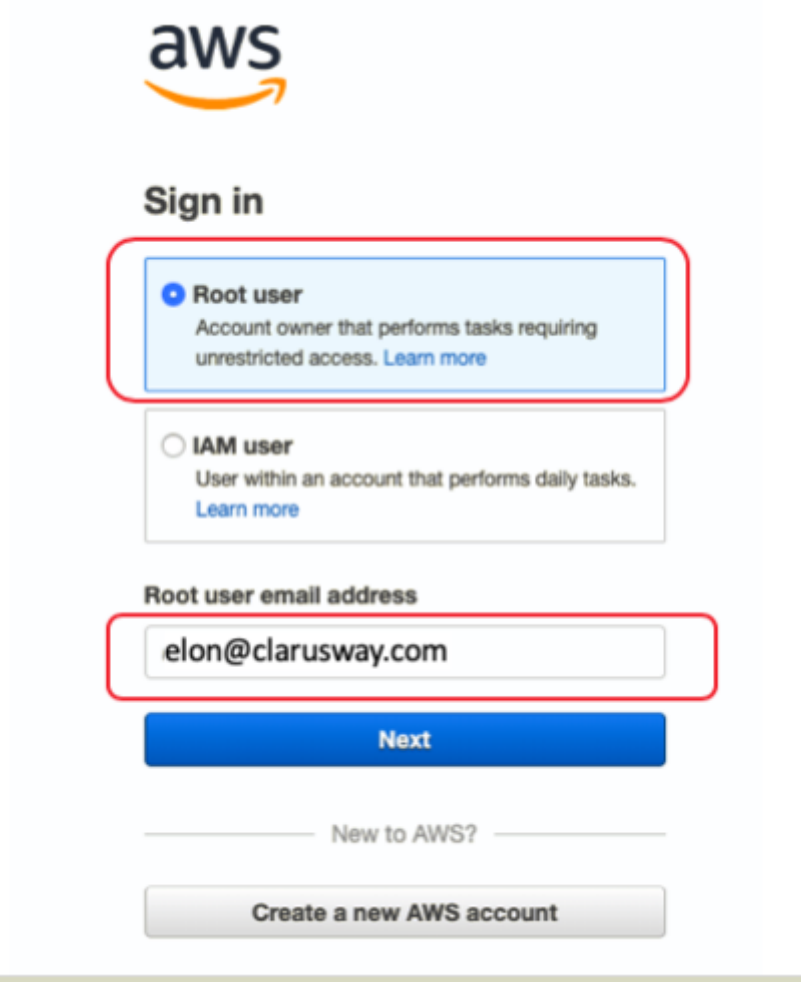
- AWS account owner is also an AWS account root user.
- An account root user has complete access to all AWS services.
- This access authorization can not be restricted in any way.
- The account owner can sign in to the AWS console as a root user by using the email address and password that was defined when creating the account. These are also known as root user credentials. MFA code can also be used as an optional but recommended feature.
- A root user can create new IAM users and give them authorization for using AWS services within the account. The limit of creating new IAM users is restricted to 5000 users per account.

- Using account root user in daily work is not recommended by AWS.
- Instead, it is recommended to create an IAM user with administrative privileges by the account owner.
- Therefore, using the root user only to create new users is considered as best practice.

Now let's connect to the account you previously created through the AWS console and take a look at the menus.

Root User Sign-in

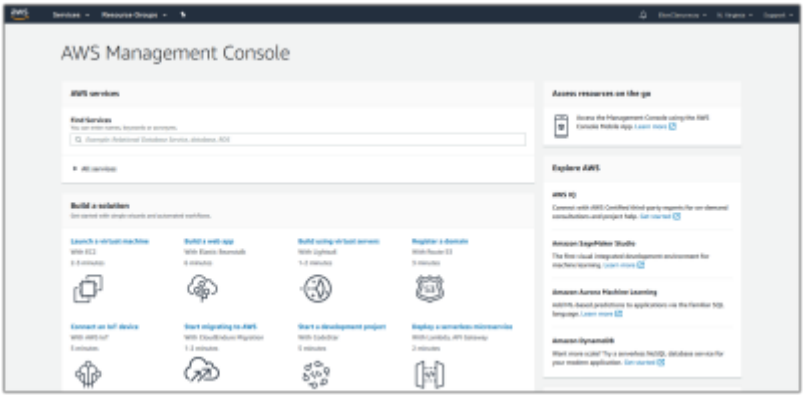
Let's connect to the AWS console by using this link.



First, choose the Root User tab. Then, type the email you defined when creating your AWS account, and then press the Next button.

Type password you defined when creating your AWS account, and then press the Next button.

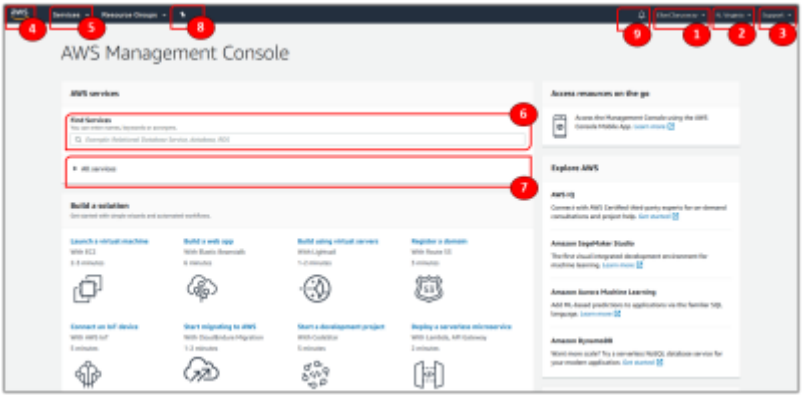
If you have activated MFA before, it will ask for MFA code at this stage, if not, the AWS console home page will open as below.



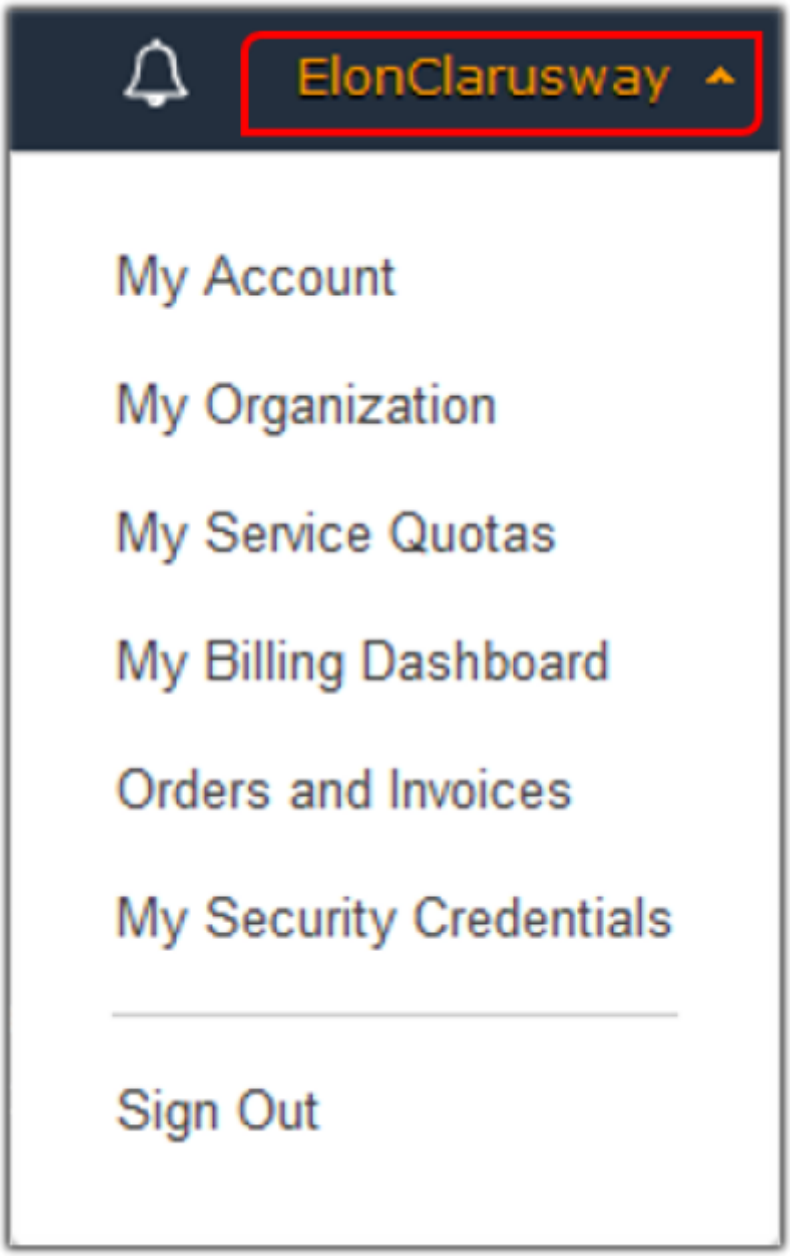
Welcome to the AWS world.

AWS Home Page Console Tour

AWS Management Console is a web application for AWS management that offers users a built-in user interface for AWS tasks. Now, let's go on a quick tour of the homepage and know some important features of the console menu.

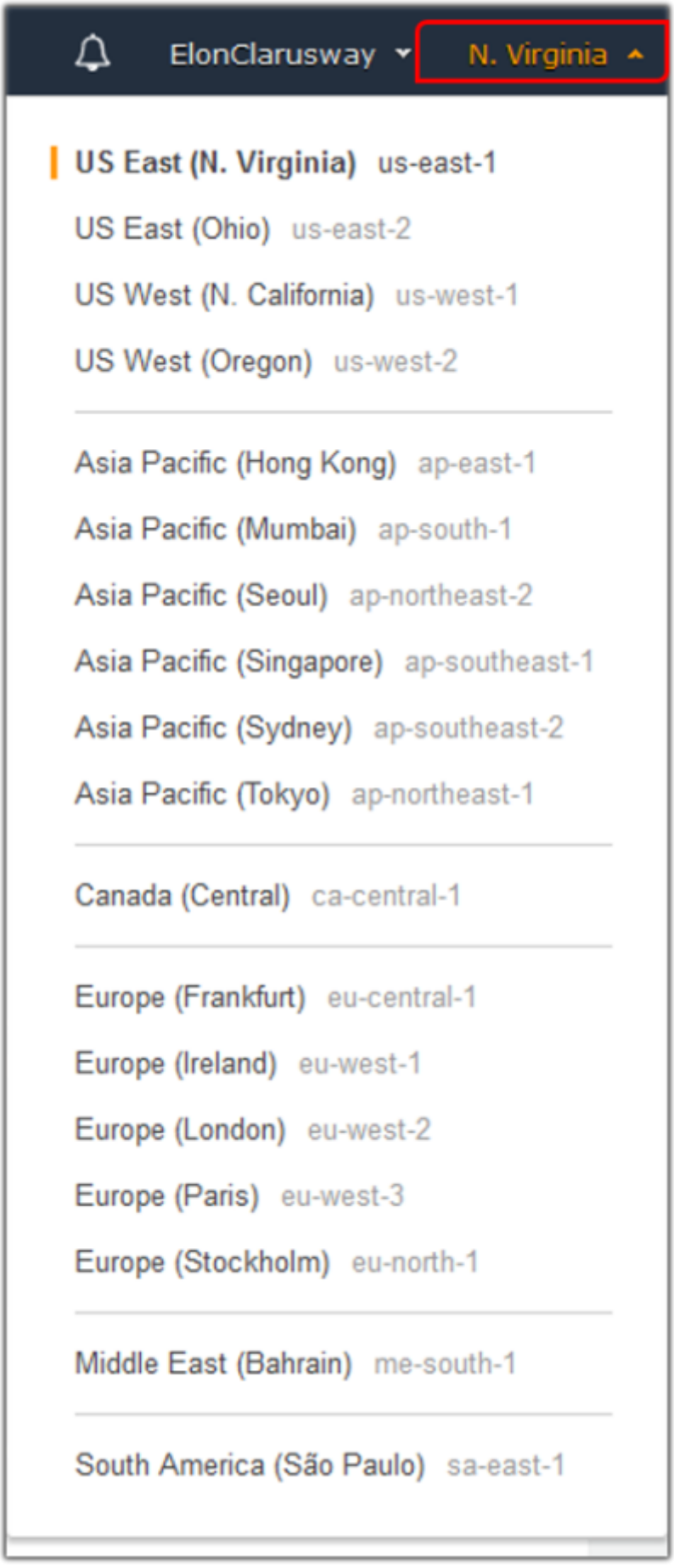


1. User Account Menu:



This menu allows you to manage all your user transactions and has been named as your user name. You can access and manage many processes such as account details, organization, billing and passwords through this menu.

2. Region:

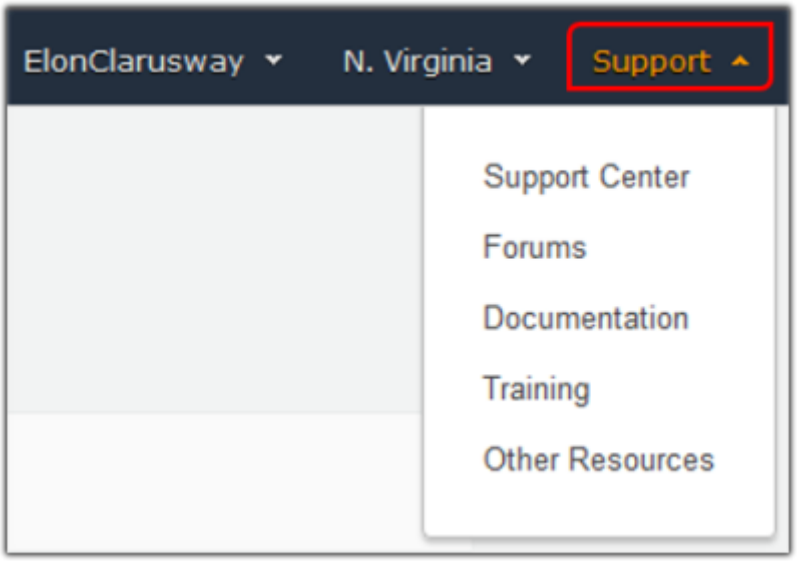


This menu shows the AWS region under operation and we can change, select and switch the region by using region list through this menu.

- After May 17, 2017, the default region when you access a resource from the AWS Management Console is US East (Ohio) (us-east-2).
- In AWS, not every region supports every service and other features of these services. You can reach which regions support which services from AWS Region Table.

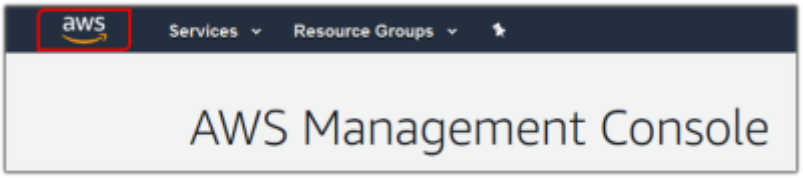
Some services such as S3 and IAM are not region-based and are therefore exhibited globally.

3. Support Menu:



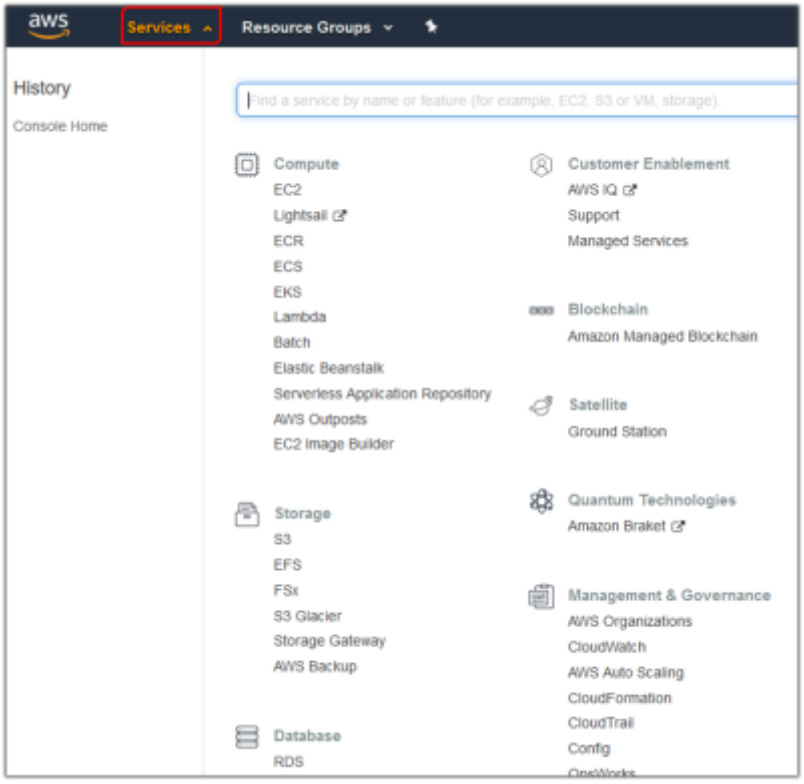
This is the menu where you can access support services of AWS such as technical, documentation, forums.

4. Home Page - Tab:



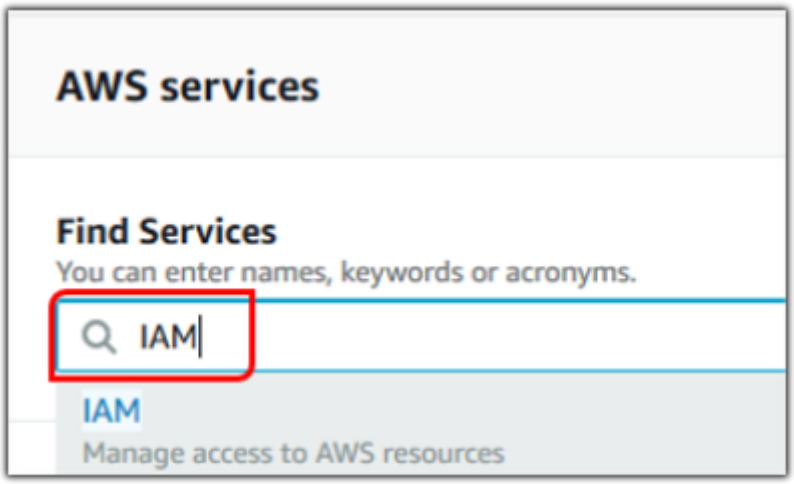
When we log into our AWS account using user credentials, the page that opens is the home page. Whenever we want to return to this home page, we can click on this tab with AWS in the upper left corner of the management console.

5. Services - Tab:



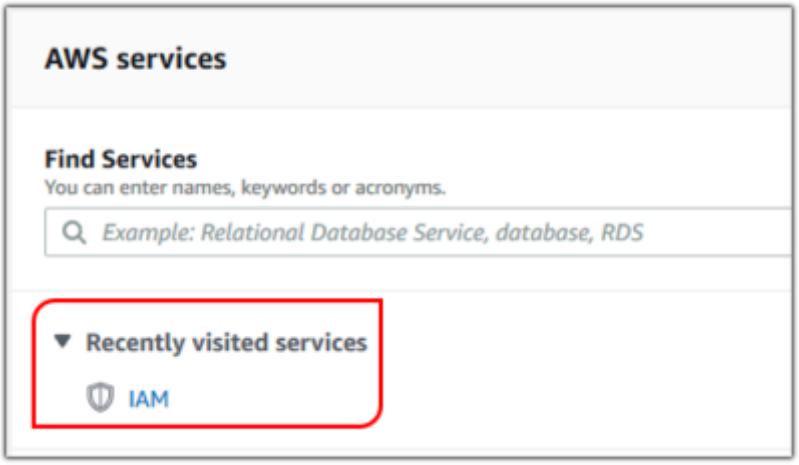
When we click the Services tab, all services offered by AWS are displayed on the management console.

6. Find Services - Search Bar:



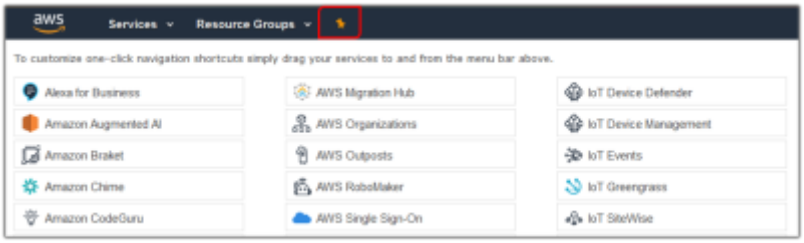
In addition to the services tab, when we want to access a service via the management console, AWS also provides us with a search bar where we can call the service we want.

7. Recently Visited Services:



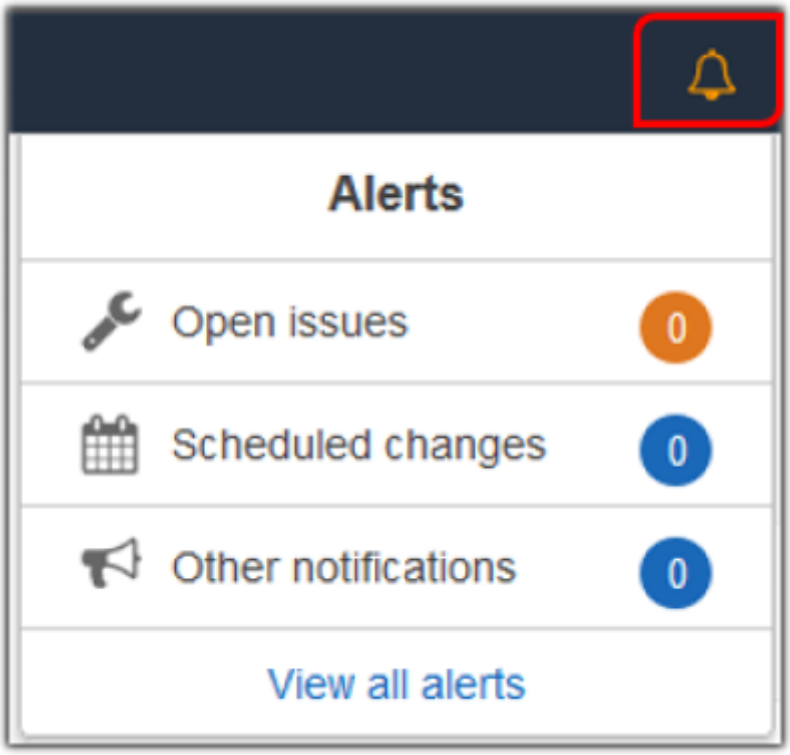
After visiting any service in the management console, this tab stores it as a recently visited service. In this way, an additional opportunity is provided by AWS to switch between services while on the home page and to reach the services you use last.

8. Pin - Menu:



With this tab, you can pin the services you use most frequently to the menu bar. In this way, you will have the opportunity to access the services, you frequently switch to, faster via the menu bar. But this is a browser-based option. When you log into your account from different browsers, if you don't pin frequently used services again, you won't be able to see them in the menu bar.

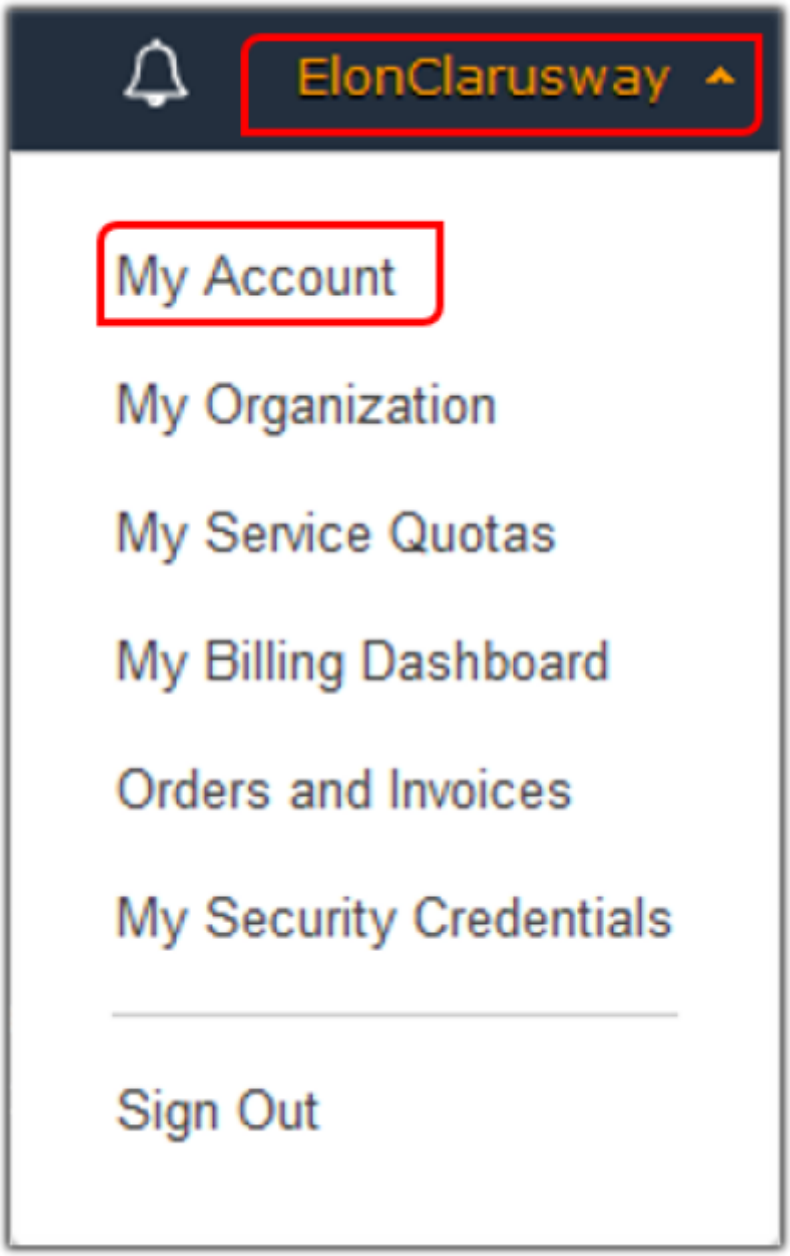
9. Notifications:



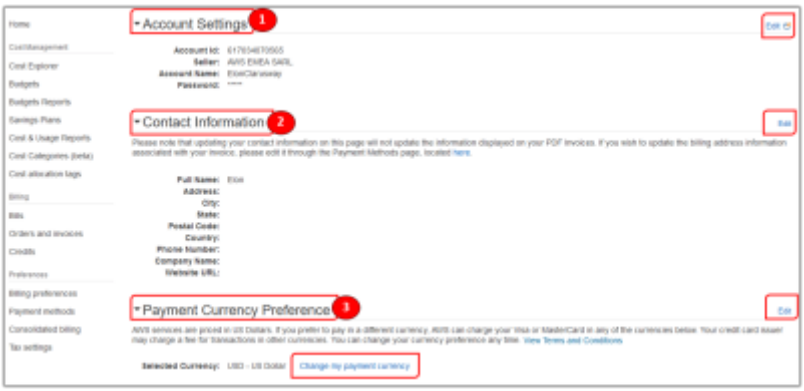
It is the tab that shows all warnings and alarms to the user such as open issues and changes. You can also view all the event logs via “View all alerts” option.

Account Information

Now let's take a look at some of our account introductory information. Click on your user name on the upper right.



After you select the My Account tab from the User Menu, the following page containing your account information will open.



1. Account Settings:

- Account Name and Account ID information are available in the Account Settings section.

- You can view and rearrange your “name, email and password” at any time by clicking the Edit tab on the right.

```
  Account ID is a unique 12-digit number like 123456789000 and can not be
changed. It's automatically created by AWS when you had created your
account. Account ID helps to distinguish your resources from other AWS
accounts resources.
```

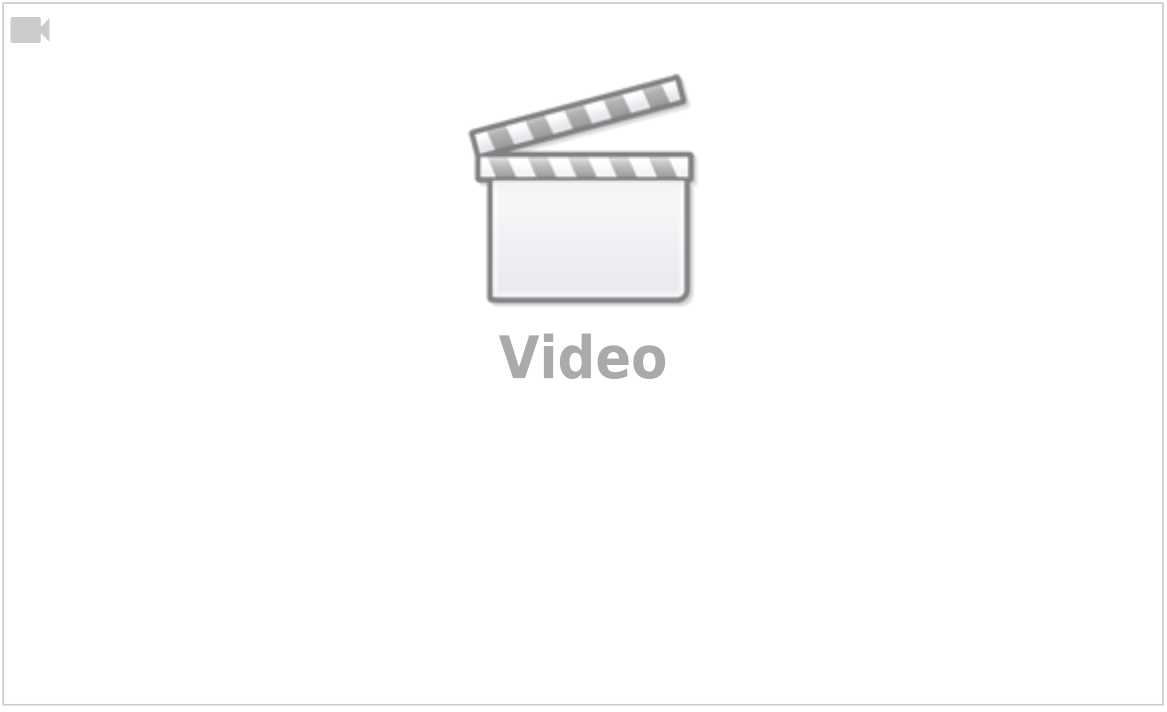
2. Contact Information:

- In the Contact Information section, your personal information such as name, address, phone number, website, and company name are included.
- You can also update your information in this section at any time by using the Edit tab.
- The point to be noted here is that; when you update your contact information, the information displayed on your PDF invoices will not be updated automatically. You should also edit it through the Payment Methods page.

3. Payment Currency Preference:

- AWS services are priced in US Dollars.
- If you prefer to pay in a different currency, AWS can charge your Visa or MasterCard in any of the currencies.
- You can change your currency preference at any time via Change my payment currency tab.

Creating IAM User



- Access Type:

AWS offers us 2 options as “Access Type” for the user. Programmatic access, AWS Management Console access. Here, we determine how the user we created can access AWS resources.

If AWS Management Console access is selected, the user can log in to AWS via the web browser and perform their operations through the AWS console. To access the console, the user needs a

“username and password”.

Additionally, if we also select Programmatic access, this user can also access AWS resources outside the console via AWS CLI (Command Line Interface), AWS API's (Application Programming Interface) and AWS SDK's (Software Development Kit). In order to access AWS resources, the user must have additional information called Access Key and Secret Access Key.

- Credentials:

The created user can access the AWS console with the username and password via the URL in the video.

Since we also define the Programmatic Access authorization for the user, the Access key ID and Secret access key information of the user has also been created. It is useful to save this information elsewhere or to download it by clicking the Download .csv option. Because this information appears here once and we cannot reach this information through the console again. However, if we want, we can recreate these keys later.

MFA Activation

AWS offers MultiFactor Authentication (MFA) to sign in to the AWS Management Console. Users not only need to provide a password or access key to work with your account, but also a code from a device that has been configured specifically.

The video below shows how to activate MFA for Root User but you can perform IAM user MFA activation similar to root user MFA activation as described in the video.