**IMPORTANCE OF CLOUD COMPUTING IN IT BUSINESS**

CSIW ZG628T DISSERTATION (Mid Submission)

by

Sapna Badola

2019wa86057

Dissertation Work carried out at

Wipro Technologies, Pune

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

Pilani (Rajasthan) India

September 2023

**CSIW ZG628T DISSERTATION**

IMPORTANCE OF CLOUD COMPUTING IN IT BUSINESS

Submitted in partial fulfilment of the requirements of

M. Tech Computing Systems & Infrastructure Degree Program

by

Sapna Badola

2019wa86057

Under the supervision of

Vaibhav Modak

Dissertation work carried out at

Wipro Technologies, Pune

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

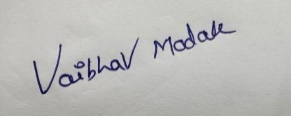
PILANI (RAJASTHAN)

(September 2023)

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**CERTIFICATE**

This is to certify that the Dissertation entitled **IMPORTANCE OF CLOUD COMPUTING IN IT BUSINESS** and submitted by **Sapna Badola** ID No. **2019wa86057** in partial fulfillment of the requirements of CSIW ZG628T Dissertation, embodies the work done by him/her under my supervision.



Signature of the Supervisor

Date: 07th Sep 2023 Name: Vaibhav Modak

Designation: Project Manager

***FORM BITS – D – 1***

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**First Semester 2023- 24**

**CSIW ZG628T DISSERTATION**

**MID SEMESTER EVALUATION FORM**

**Section I**

(To be filled by the student and returned to the Supervisor)

ID No. 2019wa86057 Name of Student: Sapna Badola

Name of Supervisor: Vaibhav Modak

Name of the Examiner(s): Rakesh Mohanty and Debashree Sensharma

Dissertation Title: IMPORTANCE OF CLOUD COMPUTING IN IT BUSINESS

**Section II**

(To be filled by the Supervisor in consultation with the examiner(s))

**Comments on the dissertation from Examiner and Supervisor (Select Y or N)**

1. **Quantum of work**
   1. **Justifiable as efforts for 8 weeks duration Y**
   2. **Work is in line with the commitments made in outline Y**
2. **Type of work**
   1. **Client assignment Y**
   2. **Organization specific task Y**
   3. **General study project such as white paper N**
   4. **Any other (kindly elaborate below in a line or two if Y N**
3. **Nature of work**
   1. **Routine in nature Y**
   2. **Involved creativity and rational thinking Y**

**Kindly elaborate below if answer for above is “Y”**

**Supervisor :** Nowadays, everything is migrating to cloud, hence it’s very important for everyone to have a basic understanding of cloud infrastructure**.** Quality of data is very good, topics well chosen.

1. **Evaluation methodology**
   1. **Evaluation done based on presentation to supervisor and examiner Y**
   2. **Evaluation done through Viva conducted by supervisor and examiner Y**
   3. **Student regularly interacted with supervisor and incorporated the suggestions made Y**
   4. **Brief description on the report submitted, quality of presentation and suggestions given for improvement**

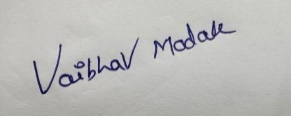
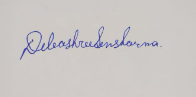
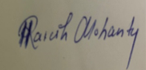
Additional sheet may be used if more space is required

**Supervisor:** The report has the detailed information about the project. Excellent presentation.

**Examiner:** Project has detailed information. Excellent requirement gathering.

1. **Mid semester evaluation matrix**

**Tick the appropriate box (1 is lowest and 5 is the highest)**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dimension Rank🡪** | **1** | **2** | **3** | **4** | **5** |
| **Student abilities in general** | | | | | |
| Understanding of the subject of dissertation |  |  |  |  | **Checkmark with solid fill** |
| Creative thinking ability to come up with new ideas |  |  |  |  | **Checkmark with solid fill** |
| **Viva / Seminar presentation** | | | | | |
| Communication ability |  |  |  |  | **Checkmark with solid fill** |
| Organization of material |  |  |  |  | **Checkmark with solid fill** |
| Response to review questions |  |  |  |  | **Checkmark with solid fill** |
| Cohesive thinking ability |  |  |  |  | **Checkmark with solid fill** |
| **Report submitted** | | | | | |
| Report structure and format |  |  |  |  | **Checkmark with solid fill** |
| Technical content of the report |  |  |  |  | **Checkmark with solid fill** |
| Explanation on the significance of the assignment |  |  |  |  | **Checkmark with solid fill** |
| Analysis of alternative approaches |  |  |  |  | **Checkmark with solid fill** |

Date: 07th Sep 2023 Signature of examiner(s) Signature of Supervisor

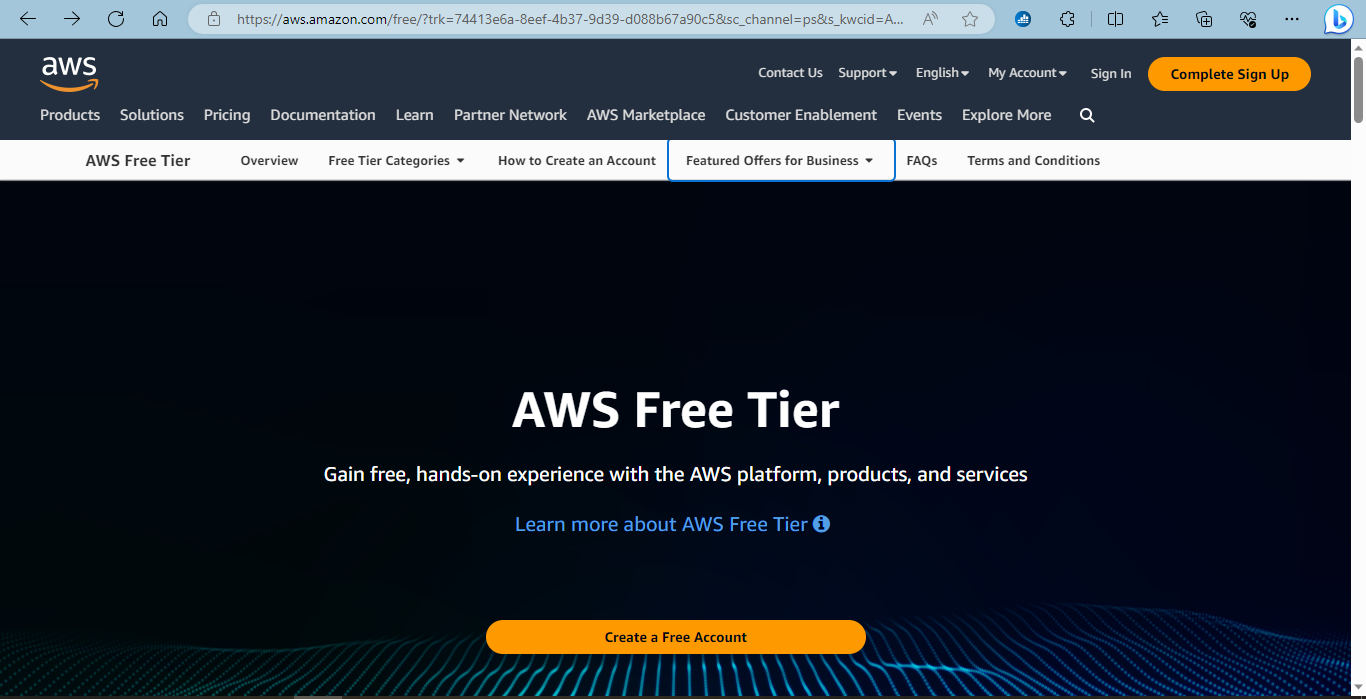
**Note: Additional paper can be used for including further comments that is relevant to the work, if required****TOPICS (Mid Sem)**

1. [Creation of AWS Account (Free Tier) for practical demonstration.](#Creation_of_AWS_Account)
2. [What is Cloud Computing, detailed information](#What_is_Cloud_Computing)
3. [Benefits of Cloud Computing in IT infrastructure](#Benefits_of_Cloud_Computing)
4. [What are the services offered by AWS, how we can use them in our IT industry.](#Services_offered_by_AWS)
5. [Client-server model, how it is beneficial for us](#Fifth)
6. [Cloud Computing Deployment models and usage](#sixth)
7. [Introduction to Amazon EC2](#seventh)
8. [EC2 instance types and benefits in our IT industry](#eighth)
9. [Amazon EC2 pricing, how we can use them for our development](#nineth)

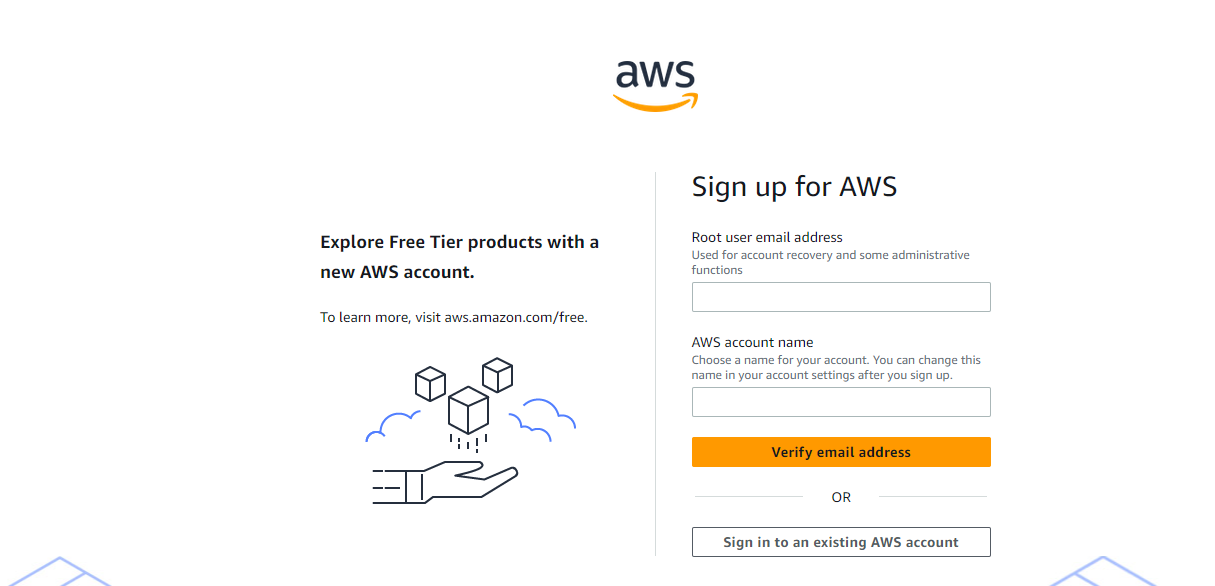
10. [Elastic Load Balancing (ELB)](#tenth)

1. [Messaging and Queuing](#eleven)
2. [Amazon SQS and SNS](#twelve)
3. [Additional Compute services](#thirteen)
4. [Amazon ECS and EKS](#forteen)
5. [AWS Fargate](#fifteen)
6. [AWS Lambda](#sixteen)
7. [Regions, how to determine correct regions](#seventeen)
8. [Availability Zones](#eighteen)
9. [Edge locations and benefits](#nineteen)
10. [Amazon CloudFront](#twenty)
11. [References/Bibliography](#references)

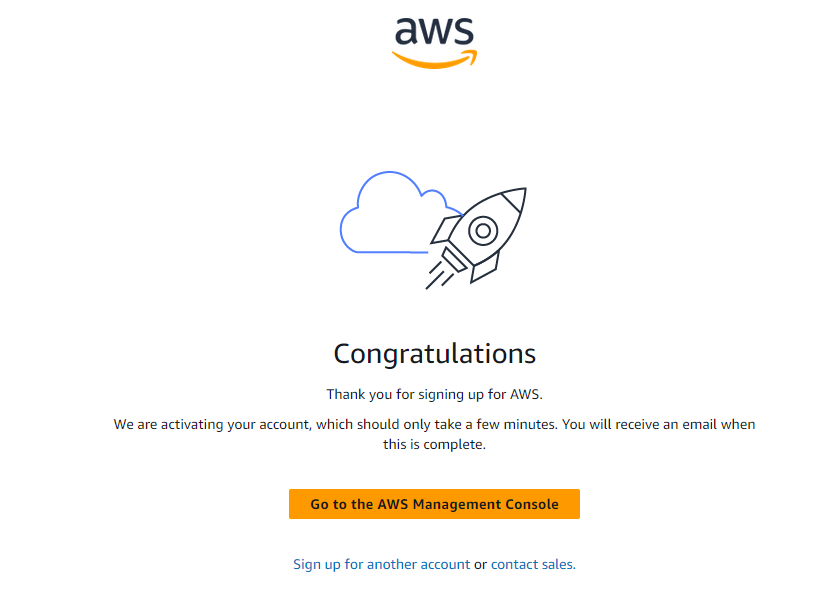
**1. Creation of AWS Account (Free Tier) for practical demonstration.**



First, we need to login into AWS site for account creation.

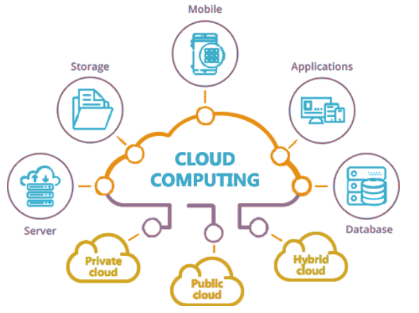


Enter all the information for creating your AWS account.



Now, after creating the AWS account we can work on this free tier for our practical demonstration.

**2. What is Cloud Computing?**

Nowadays, **Cloud computing** is adopted by every company, whether it is a MNC or a startup and many are still migrating towards it because of the cost-cutting, lesser maintenance, and the increased capacity of the data with the help of servers maintained by the cloud providers. One more reason for this drastic change from the On-premises servers of the companies to the Cloud providers is the**‘Pay as you go**’ service provided by them i.e., you only must pay for the service which you are using. The disadvantage On-premises server holds are that if the server is not in use the company still must pay for it.

Cloud computing means storing and accessing the data and programs on remote servers that are hosted on the internet instead of the computer’s hard drive or local server. Cloud computing is also referred to as Internet-based computing, it is a technology where the resource is provided as a service through the Internet to the user. The data which is stored can be files, images, documents, or any other storable document.

Some operations which can be performed with cloud computing are –

Storage, backup, and recovery of data

Delivery of software on demand

Development of new applications and services

Streaming videos and audio

**Why Cloud Computing?**

Here are the top reasons why to switch to Cloud Computing instead of owning a database server.

**1. Reduces cost :**

The cost-cutting ability of businesses that utilize cloud computing over time is one of the main advantages of this technology. On average 15% of the total cost can be saved by companies if they migrate to the cloud. Using cloud servers businesses will save and reduce costs with no need to employ a staff of technical support personnel to address server issues. There are many great business modules regarding the cost-cutting benefits of cloud servers such as the Coca-Cola and Pinterest case studies.

**2. More storage :**

For software and applications to execute as quickly and efficiently as possible, it provides more servers, storage space, and computing power. Many tools are available for cloud storage such as Dropbox, OneDrive, Google Drive, iCloud Drive, etc.

**3. Employees using cloud computing have better work-life balance :**

Direct connections between cloud computing benefits, and the work and personal lives of an enterprise’s workers can both improve because of cloud computing. Even on holidays, the employees must work with the server for its security, maintenance, and proper functionality. But with cloud storage the thing is not the same, employees get ample of time for their personal life and the workload is even less comparatively.

**Characteristics of Cloud Computing:**

**Scalability:** With Cloud hosting, it is easy to grow and shrink the number and size of servers based on the need. This is done by either increasing or decreasing the resources in the cloud. This ability to alter plans due to fluctuations in business size and needs is a superb benefit of cloud computing, especially when experiencing a sudden growth in demand.

**Instant:** Whatever you want is instantly available in the cloud.

**Save Money:** An advantage of cloud computing is the reduction in hardware costs. Instead of purchasing in-house equipment, hardware needs are left to the vendor. For companies that are growing rapidly, new hardware can be large, expensive, and inconvenient. Cloud computing alleviates these issues because resources can be acquired quickly and easily. Even better, the cost of repairing or replacing equipment is passed to the vendors. Along with purchase costs, off-site hardware cuts internal power costs and saves space. Large data centers can take up precious office space and produce a large amount of heat. Moving to cloud applications or storage can help maximize space and significantly cut energy expenditures.

**Reliability:** Rather than being hosted on one single instance of a physical server, hosting is delivered on a virtual partition that draws its resource, such as disk space, from an extensive network of underlying physical servers. If one server goes offline it will have no effect on availability, as the virtual servers will continue to pull resources from the remaining network of servers.

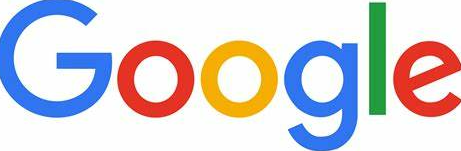
**Physical Security:** The underlying physical servers are still housed within data centers and so benefit from the security measures that those facilities implement to prevent people from accessing or disrupting them on-site.

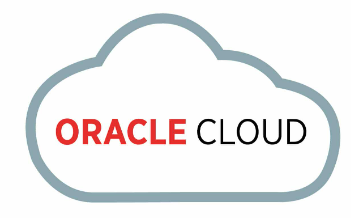
**Outsource Management:** When you are managing the business, someone else manages your computing infrastructure. You do not need to worry about management as well as degradation.

**Top leading Cloud Computing companies**

**Amazon Web Services(AWS):** One of the most successful cloud-based businesses is Amazon Web Services(AWS), which is an Infrastructure as a Service(Iaas) offering that pays rent for virtual computers on Amazon’s infrastructure.

**Microsoft Azure Platform:** Microsoft is creating the Azure platform which enables the .NET Framework Application to run over the internet as an alternative platform for Microsoft developers. This is the classic Platform as a Service(PaaS).

**Google:** Google has built a worldwide network of data centers to service its search engine. From this service, Google has captured the world’s advertising revenue. By using that revenue, Google offers free software to users based on infrastructure. This is called Software as a Service(SaaS).

**IBM Cloud** is a collection of cloud computing services for businesses provided by the IBM Corporation. It provides infrastructure as a service, software as a service, and platform as a service.

**Oracle Cloud** is a collection of cloud services offered by Oracle Corporation, including infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

**Alibaba Cloud** is the cloud computing arm of Alibaba Group, providing a comprehensive suite of global cloud computing services to power both their international customers’ online businesses and Alibaba Group’s own e-commerce ecosystem.

**Tencent Cloud** is a cloud service platform provided by Tencent. It provides a range of services such as virtual machines, storage, databases, and analytics.

**Rackspace** is a provider of hybrid cloud computing, founded in 1998. It provides managed hosting, cloud hosting, and email and app services.

**Salesforce** – A cloud-based customer relationship management (CRM) platform used for sales, marketing, and customer service.

**VMware Cloud** – A cloud platform by VMware, offering services such as virtualization, cloud management, and network virtualization.

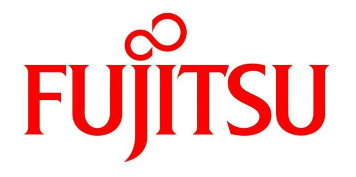
**DigitalOcean** – A cloud platform focused on providing easy-to-use, scalable computing services.

**Red Hat OpenShift** – A cloud platform by Red Hat, offering container-based application development and management.

**Cisco Cloud** – A cloud platform by Cisco, offering a range of services including networking, security, and application development.

**HP Helion** – A cloud platform by HP, offering services such as computing, storage, and networking.

**SAP Cloud Platform** – A cloud platform by SAP, offering services such as analytics, application development, and integration.

**Fujitsu Cloud** – A cloud platform by Fujitsu, offering services such as computing, storage, and networking.

**3. Benefits of Cloud Computing in IT infrastructure:**

**High availability**

A highly available system is one that experiences negligible downtime. Downtime is typically counted in seconds rather than minutes or hours since cloud-based services rarely go down. Common causes of downtime in an on-premises data center include the following:

power outages

natural disasters

hardware failures

understaffed IT departments

sabotage

**Reliability**

Reliability describes how well a service performs the tasks it promises to do. It ensures highly available databases don't randomly corrupt records or delete messages. Cloud providers routinely upgrade, update, patch and test their systems to make sure their services perform as promised. They further guarantee the reliability of their services in SLAs.

**Scalability**

What happens with an on-premises workload when demand outstrips capacity? To scale an on-premises data center, you would need to buy additional servers, install more CPUs, add memory to existing systems, expand the network and hope your upgraded infrastructure keeps pace with demand. Taking these steps is costly, time-consuming and error prone.

If you need more processing power, you can add more virtual CPUs to your EC2 instances on AWS. Or, just add virtual RAM to your ECS instances on Alibaba. Also, if your Kubernetes cluster needs more throughput, you can add new replica sets via a few clicks.

In the cloud, you can scale your architecture in minutes and with the click of a button.

**Elasticity**

The extra hardware you scale up can meet a temporary spike in demand, but what happens when demand trails off? You can scale cloud-based services as needed.

**Agility**

Productive developers need to experiment with new software and test their changes against various server configurations. This can be time-consuming, even for the most experienced developer. In the cloud, it takes only seconds for a developer to start an IBM Virtual Server or a DigitalOcean Droplet that runs a fully configured application stack.

One of the cloud computing benefits developers loves is that it frees them from the time-consuming chore of managing infrastructure.

**Cost savings**

In the cloud, capacity planning is no longer guesswork. You simply scale up and down as needed. You don't have to spend millions of dollars up front for software licenses or mainframe servers. And you'll never run into the problem of having bought too much hardware. With autoscaling, you always have a rightsized environment.

Also, you only pay for what you use, as you use it. Since there's no big, upfront expenditures to make, your costs become operational expenses. Also, because of the cost efficiencies that come with the cloud's economics of scale, costs are often lower than what you could achieve by running an on-premises data center of your own.

**Global reach**

To reduce application latency, a data center should reside near its users.

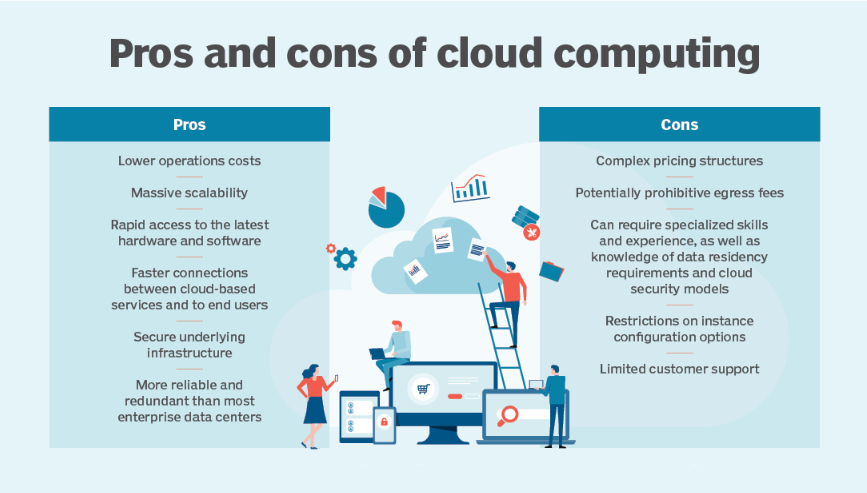
AWS and Azure have data centers located on six of the seven continents; Google and IBM are on five. That immediate global reach is one of the most compelling benefits of cloud computing, especially for organizations that service customers around the globe.

With cloud-based services, you can deploy applications into any region on the globe. You can also use edge locations around the world that have the power to cache data and further reduce application latency.

Achieving this type of global reach on your own would be incredibly difficult and prohibitively expensive. In the cloud, worldwide deployment of your applications is instant and relatively inexpensive.

**Enhanced security**

Some detractors suggest that moving data and applications to the cloud creates a security risk, but that is not the case.

Take AWS, for example: All data that flows across the AWS global network is automatically encrypted. Most AWS services, such as S3, provide the option to encrypt all data at rest, so that if a data storage device is compromised, the information on it is indecipherable.

Top cloud vendors provide many built-in tools to monitor for security noncompliance. For example, AWS Config, Google Cloud Asset Inventory and Azure Security Control monitor assets across projects and can complete compliance checks.

Built-in encryption options, mandated encryption between data centers, and the various tools that help you track user changes and identify noncompliant configurations are not available out of the box in an on-premises data center.

**Automation**

Every AWS, Azure, Google Cloud and IBM component comes with an API interface that makes it fully programmable. Developers can create, configure, query and destroy cloud-based resources with SDKs written in Java, Python, JavaScript and C++.

This built-in API interface allows developers to execute the following:

fully code around the provisioning of infrastructure;

program around mundane, manual tasks; and

automate complex, high-risk, error-prone tasks.

With the cloud, you can automate difficult tasks that could threaten the sanctity of your data center when performed improperly.

**4. What are the services offered by AWS, how we can use them in our IT industry?**

Below are the top 10 AWS Services List. As businesses are switching towards AWS for growth, it’s also mandatory to look after the services that can be used in different scenarios. AWS provides immense features or services which you could use to enhance productivity and get a fruitful solution. You can use any of them as per your project or business needs.

**1. Amazon EC2 (Elastic Cloud Compute)**

Amazon EC2 is the fastest cloud computing service provided by AWS. It offers virtual, secure, reliable, and resizable servers for any workload. Through this service, it becomes easy for developers to access resources and also facilitates web-scale cloud computing. This comes with the best suitable processors, networking facilities, and storage systems. Developers can quickly and dynamically scale capacities as per business needs. It has over 500 instances and you can also choose the latest processor, operating system, storage, and networking to help you choose according to the needs of the business. Also, with Amazon EC2, you only must pay for what you use, and also as per the time period, scale with amazon EC2 auto-scaling has optimal storage and can optimize CPU configurations.

**2. Amazon RDS (Relational Database Services)**

Amazon RDS (Relational Database Service) is another service provided by AWS which is a managed database for PostgreSQL, MariaDB, MySQL, and Oracle. Using Amazon RDS, you can set up, operate, and scale databases in the cloud. It provides high performance by automating the tasks like database setup, hardware provisioning, patching, and backups. Also, it helps in cost optimization by providing high availability, compatibility, and security for resources, and there’s no need to install and manage the database software. during its usage. As per the need, you can easily choose any engine out of 15+ engines some of them being MySQL, PostgreSQL, Oracle, etc. It is a highly secure and easily available AWS service.

**3. Amazon S3 (Simple Storage Service)**

With Amazon, it has become easy to store data anytime, anywhere. Amazon S3 (Simple Storage Service), one of the best services provided by AWS is an object storage service offering scalability, availability, security, and high performing. You can also retrieve data, data here is stored in “storage classes” where there’s no requirement of extra investment and you can also manage it well. Amazon S3 is the perfect fit for big businesses where a large amount of data is managed for varied purposes. It comes with handling any volume of data with its robust access controls, and replication tools prevent accidental deletion, and maintains data version controls.

**4. Amazon IAM (Identity and Access Management)**

Amazon IAM (Identity and Access Management) allows users to securely access and manage resources. To achieve complete access to the tools and resources provided by AWS, AWS IAM is the best AWS service. It gives you the right to have control over who has authorization (signed in) and authentication (has permissions) access to the resources. It comes with attribute-based access control which helps you to create separate permissions based on the user’s attributes such as job role, department, etc. Through this, you can allow or deny access given to users. AWS IAM has complete access or is a central manager for refining permissions across AWS. He/She handles who can access what.

**5. Amazon EBS (Elastic Block Store)**

Amazon EBS is the next service provided by AWS which is a block storage solution specifically designed for Amazon EC2. Throughout a workload of any size, Amazon EBS helps to securely manage transactions. You can handle diverse workloads, be it relational, non-relational, or business applications. You get to choose between five different volume types so as to achieve effectiveness and optimum cost. It helps to resize workloads for big data analytics engines such as Hadoop and Spark. Its lifecycle management creates policies to create and manage backups effectively. It supports high-performance scaling workloads such as Microsoft, and SAP products.

**6. Amazon Lambda**

Another promising service by AWS is Amazon Lambda which is a serverless and event-driven computing service that lets you run code for virtual applications or backend services automatically. You need to worry about servers and clusters when working with solutions using Amazon Lambda. It is also cost-effective where you must only pay for the services you use. As a user, your responsibility is to just upload the code and Lambda handles the rest. Using Lambda, you get precise software scaling and extensive availability. With hundreds to thousands of workloads per second, AWS Lambda responsibly handles code execution requests. It is one of the best services provided by AWS for developers.

**7. Amazon EFS (Elastic File System)**

Amazon EFS (Elastic File System) is a simple and serverless system where you can create and configure file systems without provisioning, deploying, patching, and maintaining. It is a scalable NFS file system made for use in AWS cloud services and on-premises resources. Also, it has no minimum fee or setup charge. You pay for the storage you use such as –

for provisioned throughput

automatically expand and shrink as per the addition and removal of files

read and write access to data stored in Infrequent Access storage classes

It is a scalable service where you can scale up to petabytes without thinking about the performance of the application.

**8. Amazon CloudFront**

Amazon CloudFront is an AWS service for content delivery networks, it delivers content globally, offering high performance and security and also at high transfer speeds and low latency(rate of time). It uses automated network mapping and intelligent routing mechanisms for delivering content to the destination. It has edge locations (worldwide network of data centers) used during content delivery. Using traffic encryptions and access controls, you can also enhance the security of data. It seamlessly integrates with systems like Amazon S3, Amazon EC2, and Lambda to manage custom code. Also, there’s no additional data transfer fee when connected with Amazon S3 and Amazon EC2.

**9. Amazon SNS (Simple Notification Service)**

It is a web service provided by AWS, which is a fully managed solution for messaging having low-cost infrastructure. It is used for bulk message delivery and direct chat with the customers through system-to-system or app-to-person between decoupled microservice apps. It is used to easily set up, operate, and send notifications from the cloud. It is a messaging service between Application to Application (A2A) and Application to Person (A2Person) and sends notifications in two ways – A2A and A2P. A2P allows many-to-many messaging between microservices, distributed systems, and event-driven serverless applications, allowing you to send messages to customers with SMS texts, email, and push notifications.

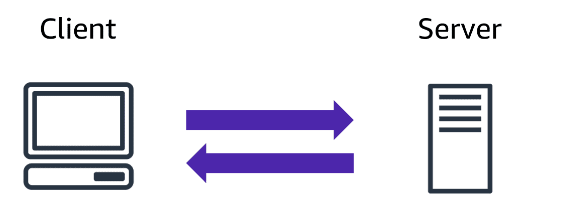
**10. Amazon VPC (Virtual Private Cloud)**

Another AWS service is Amazon VPC (Virtual Private Cloud) which is an isolated cloud resource, it enables you to set up an isolated section where you can deploy AWS resources at scale in a virtual environment. This service is responsible to control the virtual networking environment such as resource placement, security, and connectivity. Security can be improved by applying rules for outbound and inbound connections. Also, it detects anomalies in the patterns, troubleshoots network connections, prevents data leakage, and handles configuration issues. Using VPC, you get complete access to control the environment, such as choosing IP address, subset creation, and route table arrangement.

**5. Client-server model, how it is beneficial for us:**

The Client-Server model is about a client that interacts and makes requests to a computer server.

A client is the way that the person interacts with the server.



The server does tasks for the client and returns information.

The Client-server model is a distributed application structure that partitions task or workload between the providers of a resource or service, called servers, and service requesters called clients. In the client-server architecture, when the client computer sends a request for data to the server through the internet, the server accepts the requested process and deliver the data packets requested back to the client. Clients do not share any of their resources. Examples of Client-Server Model are Email, World Wide Web, etc.

**How the Client-Server Model works ?**In this article we are going to take a dive into the Client-Server model and have a look at how the Internet works via, web browsers. This article will help us in having a solid foundation of the WEB and help in working with WEB technologies with ease.

* Client: When we talk the word Client, it means to talk of a person or an organization using a particular service. Similarly in the digital world a client is a computer (Host) i.e., capable of receiving information or using a particular service from the service providers (Servers).
* Servers: Similarly, when we talk the word Servers, it means a person or medium that serves something. Similarly in this digital world a Server is a remote computer which provides information (data) or access to services.

So, it is basically the Client requesting something and the Server serving it if its present in the database.

**Advantages of Client-Server model:**

Centralized system with all data in a single place.

Cost efficient requires less maintenance cost and Data recovery is possible.

The capacity of the Client and Servers can be changed separately.

**Disadvantages of Client-Server model:**

Clients are prone to viruses, Trojans and worms if present in the Server or uploaded into the Server.

Server are prone to Denial of Service (DOS) attacks.

Data packets may be spoofed or modified during transmission.

Phishing or capturing login credentials or other useful information of the user are common and MITM(Man in the Middle) attacks are common.

**6. Cloud Computing Deployment models and usage:**

Cloud Deployment Model functions as a virtual computing environment with a deployment architecture that varies depending on the amount of data you want to store and who has access to the infrastructure.

Types of Cloud Computing Deployment Models

The cloud deployment model identifies the specific type of cloud environment based on ownership, scale, and access, as well as the cloud’s nature and purpose. The location of the servers you are utilizing and who controls them are defined by a cloud deployment model. It specifies how your cloud infrastructure will look, what you can change, and whether you will be given services or will have to create everything yourself. Relationships between the infrastructure and your users are also defined by cloud deployment types. [Several types of cloud](https://www.geeksforgeeks.org/types-of-cloud/) computing deployment models are described below.

**Public Cloud**

The public cloud makes it possible for anybody to access systems and services. The public cloud may be less secure as it is open to everyone. The public cloud is one in which cloud infrastructure services are provided over the internet to the general people or major industry groups. The infrastructure in this cloud model is owned by the entity that delivers the cloud services, not by the consumer. It is a type of cloud hosting that allows customers and users to easily access systems and services. This form of cloud computing is an excellent example of cloud hosting, in which service providers supply services to a variety of customers. In this arrangement, storage backup and retrieval services are given for free, as a subscription, or on a per-user basis. For example, Google App Engine etc.

**Private Cloud**

The private cloud deployment model is the exact opposite of the public cloud deployment model. It’s a one-on-one environment for a single user (customer). There is no need to share your hardware with anyone else. The distinction between private and public clouds is in how you handle all the hardware. It is also called the “internal cloud” & it refers to the ability to access systems and services within a given border or organization. The cloud platform is implemented in a cloud-based secure environment that is protected by powerful firewalls and under the supervision of an organization’s IT department. The private cloud gives greater flexibility of control over cloud resources.

**Hybrid Cloud**

By bridging the public and private worlds with a layer of proprietary software, hybrid cloud computing gives the best of both worlds. With a hybrid solution, you may host the app in a safe environment while taking advantage of the public cloud’s cost savings. Organizations can move data and applications between different clouds using a combination of two or more cloud deployment methods, depending on their needs.

**7. Introduction to Amazon EC2:**

Amazon Elastic Compute Cloud (Amazon EC2) provides on-demand, scalable computing capacity in the Amazon Web Services (AWS) Cloud. Using Amazon EC2 reduces hardware costs so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage. You can add capacity (scale up) to handle compute-heavy tasks, such as monthly or yearly processes, or spikes in website traffic. When usage decreases, you can reduce capacity (scale down) again.

Amazon EC2 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). For more information about PCI DSS, including how to request a copy of the AWS PCI Compliance Package.

**Features of Amazon EC2**

Amazon EC2 provides the following high-level features:

**Instances**

Virtual servers.

**Amazon Machine Images (AMIs)**

Preconfigured templates for your instances that package the components you need for your server (including the operating system and additional software).

**Instance types**

Various configurations of CPU, memory, storage, networking capacity, and graphics hardware for your instances.

**Key pairs**

Secure login information for your instances. AWS stores the public key and you store the private key in a secure place.

**Instance store volumes**

Storage volumes for temporary data that is deleted when you stop, hibernate, or terminate your instance.

**Amazon EBS volumes**

Persistent storage volumes for your data using Amazon Elastic Block Store (Amazon EBS).

**Regions, Availability Zones, Local Zones, AWS Outposts, and Wavelength Zones**

Multiple physical locations for your resources, such as instances and Amazon EBS volumes.

**Security groups**

A virtual firewall that allows you to specify the protocols, ports, and source IP ranges that can reach your instances, and the destination IP ranges to which your instances can connect.

**Elastic IP addresses**

Static IPv4 addresses for dynamic cloud computing.

**Tags**

Metadata that you can create and assign to your Amazon EC2 resources.

**Virtual private clouds (VPCs)**

Virtual networks you can create that are logically isolated from the rest of the AWS Cloud. You can optionally connect these virtual networks to your own network.

**8. EC2 instance types and benefits:**

**General Purpose:** It provides a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads. These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories.

**Compute Optimized:** Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors. Instances belonging to this category are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.

**Memory Optimized:** Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.

**Accelerated Computing:** Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating-point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs.

**Storage Optimized:** Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latencies, random I/O operations per second (IOPS) to applications.

**HPC Optimized:** High performance computing (HPC) instances are purpose built to offer the best price performance for running HPC workloads at scale on AWS. HPC instances are ideal for applications that benefit from high-performance processors such as large, complex simulations and deep learning workloads.

**9. Amazon EC2 pricing, how we can use them for our development:**

There are four ways to pay for Amazon EC2 instances: On-Demand Instances, Savings Plans, Spot Instances, and Reserved Instances (RIs).

**On-Demand Instances**

With On-Demand Instances, you pay for compute capacity per hour or per second, depending on which instances you run. No long-term commitments or up-front payments are required. You can increase or decrease your compute capacity to meet the demands of your application and only pay the speciﬁed hourly rates for the instance you use. On-Demand Instances are recommended for the following use cases:

Users who prefer the low cost and ﬂexibility of Amazon EC2 without upfront payment or long-term commitments.

Applications with short-term, spiky, or unpredictable workloads that cannot be interrupted.

Applications being developed or tested on Amazon EC2 for the ﬁrst time.

**Savings Plans**

Savings Plans is a ﬂexible pricing model that oﬀers low prices on Amazon EC2, SageMaker, Lambda, and Fargate usage, in exchange for a commitment to a consistent amount of usage (measured in $/hour) for a one or three-year term. Savings Plans is a ﬂexible pricing model that provides savings of up to 72 percent on your AWS compute usage. This pricing model oﬀers lower prices on Amazon EC2 instances usage, regardless of instance family, size, operating system (OS), tenancy, or AWS Region, and also applies to SageMaker, Fargate, and Lambda usage.

For workloads that have predictable and consistent usage, Savings Plans can provide signiﬁcant savings compared to On-Demand pricing. It is recommended for:

Workloads with consistent and steady-state usage.

Customers who want to use diﬀerent instance types and compute solutions across diﬀerent locations.

Customers who can make a monetary commitment to use compute services over a one or three-year term.

**Spot Instances**

Amazon EC2 Spot Instances allow you to request spare Amazon EC2 computing capacity for up to 90 percent oﬀ the On-Demand price. Spot Instances are recommended for:

Applications that have ﬂexible start and end times.

Applications that are only feasible at very low compute prices.

Users with fault-tolerant and/or stateless workloads.

Spot Instance prices are set by Amazon EC2 and adjust gradually based on long-term trends in supply and demand for Spot Instance capacity.

**RIs**

Amazon EC2 Reserved Instances provide you with a signiﬁcant discount (up to 72 percent) compared to On-Demand Instance pricing. In addition, when RIs are assigned to a speciﬁc Availability Zone, they provide a capacity reservation, giving you additional conﬁdence in your ability to launch instances when you need them.

**Per-second billing**

Per-second billing saves money and has a minimum of 60 seconds billing. It is particularly eﬀective for resources that have periods of low and high usage such as development and testing, data processing, analytics, batch processing, and gaming applications. Learn more about per-second billing.

**10. Elastic Load Balancing (ELB):**

Elastic Load Balancing automatically distributes your incoming traffic across multiple targets, such as EC2 instances, containers, and IP addresses, in one or more Availability Zones. It monitors the health of its registered targets, and routes traffic only to the healthy targets. Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic. A load balancer distributes workloads across multiple compute resources, such as virtual servers. Using a load balancer increases the availability and fault tolerance of your applications. You can add and remove compute resources from your load balancer as your needs change, without disrupting the overall flow of requests to your applications.

You can configure health checks, which monitor the health of the compute resources, so that the load balancer sends requests only to the healthy ones. You can also offload the work of encryption and decryption to your load balancer so that your compute resources can focus on their main work.

**11. Messaging and Queuing:**

Microservices can help to maintain the service if one component fails. The services can be maintained because they communicate with each other, and the components are not tightly coupled.

AWS has two services that can make this integration:

* **AWS Simple Notification Service (AWS SNS)**
* **AWS Simple Queue Service (AWS SQS)**

**12. Amazon SQS and SNS:**

**What is AWS SNS?**

SNS is a cloud service for the mass delivery of messages. It is a fully managed publish-subscribe messaging and mobile communication service. It can be event-driven, with automated services responding to triggers. Distributed systems and micro services can be decoupled with messaging between them through AWS SNS. Application-to-person messaging to users is possible with SMS, mobile push, and email.

**Cloud Queue Service - AWS SQS**

AWS Simple Queue Service is also called AWS SQS. SQS is a message queuing service. It exchanges and stores messages between software components. The service adds the messages in a queue. Users or services pick up the messages from the queue. Once processed the messages gets deleted from the queue.

**13. Additional Compute services:**

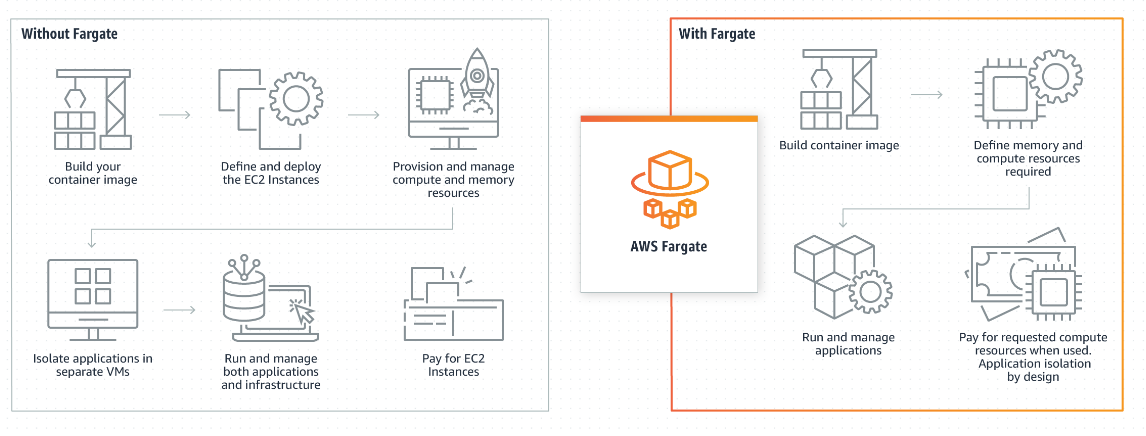
|  |  |
| --- | --- |
| **Category** | **AWS service** |
| Instances (virtual machines) | * [Amazon Elastic Compute Cloud](https://aws.amazon.com/ec2/) (Amazon EC2) — Secure and resizable compute capacity (virtual servers) in the cloud * [Amazon EC2 Spot Instances](https://aws.amazon.com/ec2/spot/)— Run fault-tolerant workloads for up to 90% off * [Amazon EC2 Auto Scaling](https://aws.amazon.com/ec2/autoscaling/) — Automatically add or remove compute capacity to meet changes in demand * [Amazon LightSail](https://aws.amazon.com/lightsail/) — Easy-to-use cloud platform that offers you everything you need to build an application or website * [AWS Batch](https://aws.amazon.com/batch/) — Fully managed batch processing at any scale |
| Containers | * [Amazon Elastic Container Service](https://aws.amazon.com/ecs/) (Amazon ECS) — Highly secure, reliable, and scalable way to run containers * [Amazon ECS Anywhere](https://aws.amazon.com/ecs/anywhere/) — Run containers on customer-managed infrastructure * [Amazon Elastic Container Registry](https://aws.amazon.com/ecr/) (Amazon ECR) — Easily store, manage, and deploy container images * [Amazon Elastic Kubernetes Service](https://aws.amazon.com/eks/) (Amazon EKS) — Fully managed Kubernetes service * [Amazon EKS Anywhere](https://aws.amazon.com/eks/) — Create and operate Kubernetes clusters on your own infrastructure * [AWS Fargate](https://aws.amazon.com/fargate/) — Serverless compute for containers * [AWS App Runner](https://aws.amazon.com/apprunner/) — Build and run containerized applications on a fully managed service |
| Serverless | * [AWS Lambda](https://aws.amazon.com/lambda/) — Run code without thinking about servers. Pay only for the compute time you consume. |
| Edge and hybrid | * [AWS Outposts](https://aws.amazon.com/outposts/) — Run AWS infrastructure and services on premises for a truly consistent hybrid experience * [AWS Snow Family](https://aws.amazon.com/snow/) — Collect and process data in rugged or disconnected edge environments * [AWS Wavelength](https://aws.amazon.com/wavelength/) — Deliver ultra-low latency application for 5G devices * [VMware Cloud on AWS](https://aws.amazon.com/vmware/) — Preferred service for all vSphere workloads to rapidly extend and migrate to the cloud * [AWS Local Zones](https://aws.amazon.com/about-aws/global-infrastructure/localzones/) — Run latency sensitive applications closer to end-users |
| Cost and capacity management | * [AWS Savings Plan](https://aws.amazon.com/savingsplans/) — Flexible pricing model that provides savings of up to 72% on AWS compute usage * [AWS Compute Optimizer](https://aws.amazon.com/compute-optimizer/) — Recommends optimal AWS compute resources for your workloads to reduce costs and improve performance * [AWS Elastic Beanstalk](https://aws.amazon.com/elasticbeanstalk/) — Easy-to-use service for deploying and scaling web applications and services * [EC2 Image Builder](https://aws.amazon.com/image-builder/) — Build and maintain secure Linux or Windows Server images * [Elastic Load Balancing](https://aws.amazon.com/elasticloadbalancing/) (ELB) — Automatically distribute incoming application traffic across multiple targets |

**14. Amazon ECS and EKS:**

**Amazon Elastic Container Service (Amazon ECS)** is a fully managed container orchestration service that simplifies your deployment, management, and scaling of containerized applications. Simply describe your application and the resources required, and Amazon ECS will launch, monitor, and scale your application across flexible compute options with automatic integrations to other supporting AWS services that your application needs. Perform system operations such as creating custom scaling and capacity rules and observe and query data from application logs and telemetry.

**Amazon Elastic Kubernetes Service (Amazon EKS)** is a managed service that eliminates the need to install, operate, and maintain your own Kubernetes control plane on Amazon Web Services (AWS). Kubernetes is an open-source system that automates the management, scaling, and deployment of containerized applications.

**15. AWS Fargate:**

AWS Fargate is a serverless, pay-as-you-go compute engine that lets you focus on building applications without managing servers. AWS Fargate is compatible with both [Amazon Elastic Container Service](https://aws.amazon.com/ecs/?pg=ln&sec=hiw) (Amazon ECS) and [Amazon Elastic Kubernetes Service](https://aws.amazon.com/eks/?pg=ln&sec=hiw) (Amazon EKS). Select any OCI-compliant container image, define memory and compute resources, and run the container with serverless compute. With multiple CPU architectures and operating systems supported; you can enjoy the benefits across a wide variety of applications.

**16. AWS Lambda:**

* Run code without provisioning or managing servers, creating workload-aware cluster scaling logic, maintaining event integrations, or managing runtimes.
* Run code for virtually any type of application or backend service. Just upload your code as a ZIP file or container image, and Lambda automatically allocates compute execution power and runs your code based on the incoming request or event, for any scale of traffic.
* Write Lambda functions in your favorite language (Node.js, Python, Go, Java, and more) and use both serverless and container tools, such as AWS SAM or Docker CLI, to build, test, and deploy your functions.

**17. Regions, how to determine correct regions?**

An AWS Region is a cluster of data centers in a specific geographic area, such as the Northeastern United States or Western Europe. It is a best practice to choose a region that is geographically close to users; this reduces latency because data reaches the users more quickly.

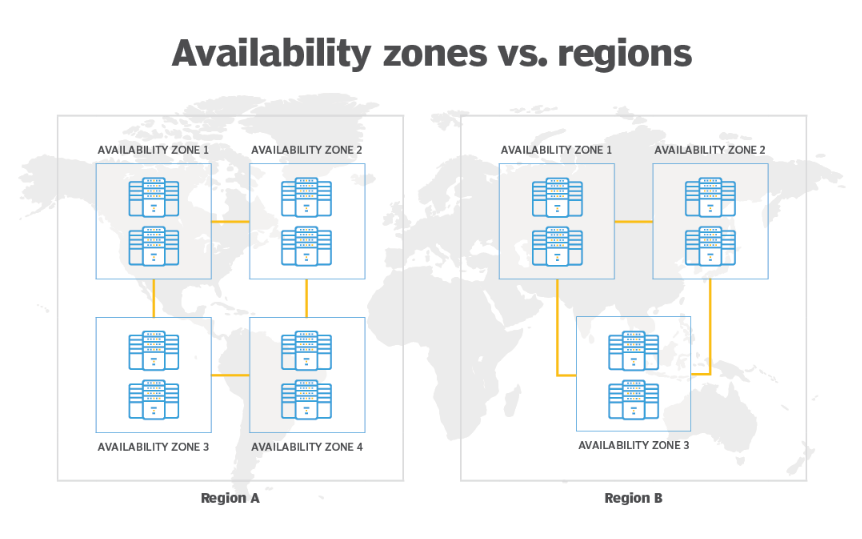
Each AWS Region includes multiple AZs. However, each AZ is restricted to a specific AWS Region. You can use multiple AZs within one Region, but you can't use the same AZ across multiple Regions. For many services, you can choose an AWS Region that specifies where your resources are managed. Regions are sets of AWS resources located in the same geographical area. You don't need to choose a Region for the [AWS Management Console](https://console.aws.amazon.com/) or for some services, such as AWS Identity and Access Management. To learn more about AWS Regions, see [Managing AWS Regions](https://docs.aws.amazon.com/general/latest/gr/rande-manage.html) in the AWS General Reference.

**To choose a Region**

* Sign in to the [AWS Management Console](https://console.aws.amazon.com/).
* [Choose a service](https://docs.aws.amazon.com/awsconsolehelpdocs/latest/gsg/start-service.html) to go to that service's console.
* In the navigation bar, choose the name of the currently displayed Region. Then choose the Region to which you want to switch.
* To choose a default Region
* In the navigation bar, choose your account name, and then choose Settings to navigate to the Unified Settings page.
* Choose Edit next to Localization and default Region.
* Select your default Region, then choose Save changes. If you do not select a default Region, the last Region you visited will be your default.

**18. Availability Zones:**

An AZ is a standalone data center or set of data centers within a Region. Each AZ operates independently, so a failure in one won't affect others. In disaster recovery plans, enterprises use multiple AZs to [increase redundancy and reliability](https://www.techtarget.com/searchcloudcomputing/tip/Stay-online-with-these-5-AWS-disaster-recovery-best-practices). AZs shouldn't be confused with [AWS Local Zones](https://www.techtarget.com/searchcloudcomputing/tip/AWS-Local-Zones-offers-low-latency-with-clear-limitations), which are extensions of a Region. Local Zones let you choose more specific geographic locations, such as Boston or Los Angeles. They are not designed to increase workload redundancy. They are valuable if your users are concentrated in a relatively small area, as they help reduce latency and meet strict compliance requirements.



**19. Edge locations and benefits:**

**Edge locations are AWS data centers designed to deliver services with the lowest latency possible.**Amazon has dozens of these data centers spread across the world. They’re closer to users than Regions or Availability Zones, often in major cities, so responses can be fast and snappy. A subset of services for which latency really matters use edge locations, including:

* [**CloudFront**](https://aws.amazon.com/cloudfront/), which uses edge locations to cache copies of the content that it serves, so the content is closer to users and can be delivered to them faster.
* [**Route 53**](https://aws.amazon.com/route53/), which serves DNS responses from edge locations, so that DNS queries that originate nearby can resolve faster (and, contrary to what you might think, is also [Amazon’s premier database](https://www.lastweekinaws.com/blog/route-53-amazons-premier-database/)).
* [**Web Application Firewall**](https://aws.amazon.com/waf/)and[**AWS Shield**](https://aws.amazon.com/shield/), which filter traffic in edge locations to stop unwanted traffic as soon as possible.

**20. Amazon CloudFront:**

Amazon CloudFront is a web service that speeds up distribution of your static and dynamic web content, such as .html, .css, .js, and image files, to your users. CloudFront delivers your content through a worldwide network of data centers called edge locations. When a user requests content that you're serving with CloudFront, the request is routed to the edge location that provides the lowest latency (time delay), so that content is delivered with the best possible performance.

* If the content is already in the edge location with the lowest latency, CloudFront delivers it immediately.
* If the content is not in that edge location, CloudFront retrieves it from an origin that you've defined—such as an Amazon S3 bucket, a MediaPackage channel, or an HTTP server (for example, a web server) that you have identified as the source for the definitive version of your content.

**References/Bibliography:**

For this project we can take reference from below sites:

1. [Free Cloud Computing Services - AWS Free Tier (amazon.com)](https://aws.amazon.com/free/?all-free-tier.sort-by=item.additionalFields.SortRank&all-free-tier.sort-order=asc&awsf.Free%20Tier%20Types=*all&awsf.Free%20Tier%20Categories=*all)
2. [Amazon Elastic Compute Cloud Documentation](https://docs.aws.amazon.com/ec2/index.html?nc2=h_ql_doc_ec2)
3. [Amazon Simple Storage Service Documentation](https://docs.aws.amazon.com/s3/index.html?nc2=h_ql_doc_s3)
4. [Welcome to AWS Documentation (amazon.com)](https://docs.aws.amazon.com/)
5. [Amazon RDS and Aurora Documentation](https://docs.aws.amazon.com/rds/?icmpid=docs_homepage_featuredsvcs)
6. [AWS Lambda Documentation (amazon.com)](https://docs.aws.amazon.com/lambda/?icmpid=docs_homepage_featuredsvcs)
7. [Amazon DynamoDB Documentation](https://docs.aws.amazon.com/dynamodb/?icmpid=docs_homepage_featuredsvcs)
8. [Amazon Virtual Private Cloud Documentation](https://docs.aws.amazon.com/vpc/?icmpid=docs_homepage_featuredsvcs)
9. [Learner dashboard - AWS Skill Builder](https://explore.skillbuilder.aws/pages/16/learner-dashboard)
10. [On Premises VS On Cloud - GeeksforGeeks](https://www.geeksforgeeks.org/on-premises-vs-on-cloud/)