

Unit 1: Introduction to Cloud and AWS.

Classic Data Center, Virtualization, Cloud and Cloud Computing, Cloud Computing Service Models, Cloud Computing Deployment Models, Service Comparison: AWS, Azure, and GCP, Amazon Web Services (AWS) and its Benefits, AWS Global Infrastructure, AWS Regions and Replication of data between the Regions, Availability Zones and High Availability, AWS Edge Location, Different Amazon Web Services.

Classic Data Center

A data center is a facility that centralizes an organization's shared IT operations and equipment for the purposes of storing, processing, and disseminating data and applications. Because they house an organization's most critical and proprietary assets, data centers are vital to the continuity of daily operations. Consequently, the security and reliability of data centers and their information are among any organization's top priorities.

Data centers are an integral part of the enterprise, designed to support business applications and provide services such as:

- Data storage, management, backup and recovery
- Productivity applications, such as email
- High-volume e-commerce transactions
- Powering online gaming communities
- Big data, machine learning and artificial intelligence

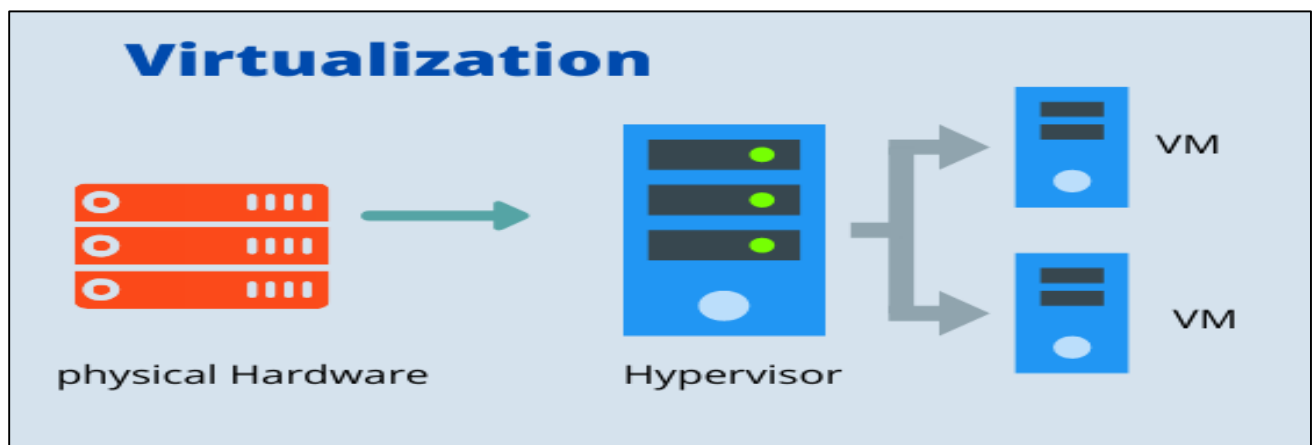
The primary elements of a data center break down as follows:

- Facility – the usable space available for IT equipment. Providing round-the-clock access to information makes data centers some of the world's most energy-consuming facilities. Design to optimize space and environmental control to keep equipment within specific temperature/humidity ranges are both emphasized.
- Core components – equipment and software for IT operations and storage of data and applications. These may include storage systems; servers; network infrastructure, such as switches and routers; and various information security elements, such as firewalls.
- Support infrastructure – equipment contributing to securely sustaining the highest availability possible. The Uptime Institute has defined four tiers of data centers, with availability ranging from 99.671% to 99.995%. Some components for supporting infrastructure include:
 - Uninterruptible Power Sources (UPS) – battery banks, generators and redundant power sources.
 - Environmental control – computer room air conditioners (CRAC); heating, ventilation and air conditioning (HVAC) systems; and exhaust systems.
 - Physical security systems – biometrics and video surveillance systems.
 - Operations staff – personnel available to monitor operations and maintain IT and infrastructure equipment around the clock.

Data centers have evolved significantly in recent years. As enterprise IT needs continue to move toward on-demand services, data center infrastructure has shifted from on-premises servers to virtualized infrastructure that supports workloads across pools of physical infrastructure and multi-cloud environments.

Virtualization

- Virtualization is the "creation of a virtual (rather than actual) version of something, such as a server, a desktop, a storage device, an operating system or network resources".
- In other words, Virtualization is a technique, which allows to share a single physical instance of a resource or an application among multiple customers and organizations.
- It does by assigning a logical name to a physical storage and providing a pointer to that physical resource when demanded.
- Virtualization is the process of running a virtual instance of a computer system in a layer abstracted from the actual hardware.
- Most commonly, it refers to running multiple operating systems on a computer system simultaneously. There are many reasons why people utilize virtualization in computing.
- To desktop users, the most common use is to be able to run applications meant for a different operating system without having to switch computers or reboot into a different system.
- For administrators of servers, virtualization also offers the ability to run different operating systems, but perhaps, more importantly, it offers a way to segment a large system into many smaller parts, allowing the server to be used more efficiently by a number of different users or applications with different needs.
- It also allows for isolation, keeping programs running inside of a virtual machine safe from the processes taking place in another virtual machine on the same host.



The virtualization process follows the steps listed below:

- Hypervisors detach the physical resources from their physical environments.
- Resources are taken and divided, as needed, from the physical environment to the various virtual environments.
- System users work with and perform computations within the virtual environment.
- Once the virtual environment is running, a user or program can send an instruction that requires extra resources from the physical environment. In response, the hypervisor relays the message to the physical system and stores the changes. This process will happen at an almost native speed.
- The virtual environment is often referred to as a guest machine or virtual machine. The VM acts like a single data file that can be transferred from one computer to another and opened in both; it is expected to perform the same way on every computer.

Cloud and Cloud Computing

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on an as-needed basis from a cloud provider.

In simple terms, cloud computing is a range of services delivered over the internet, or “the cloud.” It means using remote servers to store and access data instead of relying on local hard drives and private datacenters. Before cloud computing existed, organizations had to purchase and maintain their own servers to meet business needs.

This required buying enough server space to reduce the risk of downtime and outages, and to accommodate peak traffic volume. As a result, large amounts of server space went unused for much of the time. Today’s cloud service providers allow companies to reduce the need for onsite servers, maintenance personnel, and other costly IT resources.

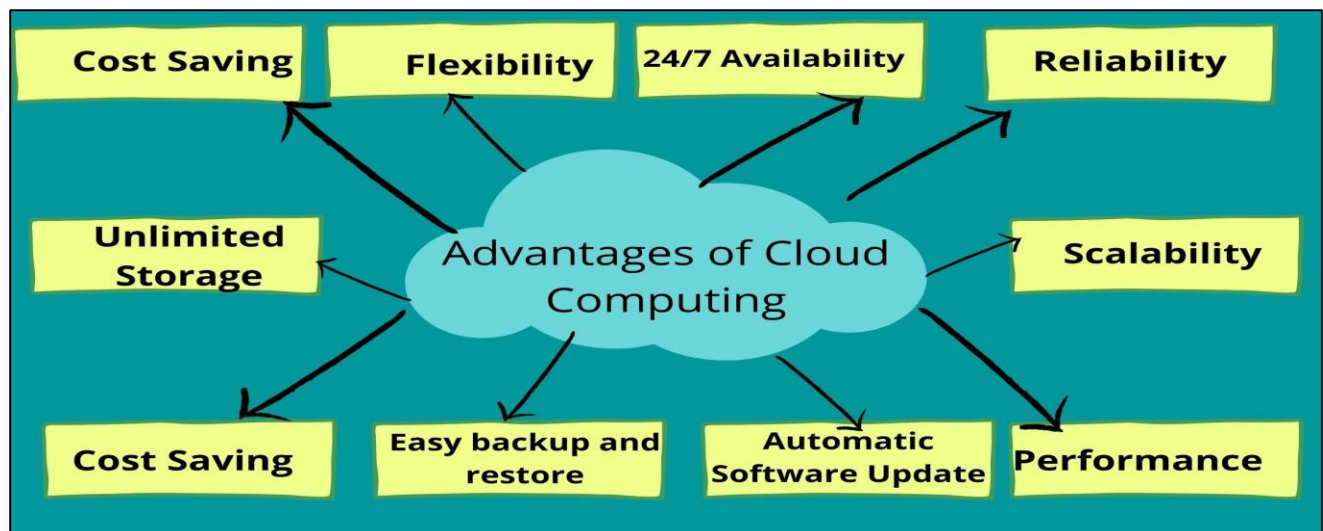
The cloud enables users to access the same files and applications from almost any device, because the computing and storage takes place on servers in a data center, instead of locally on the user device. This is why a user can log in to their Instagram account on a new phone after their old phone breaks and still find their old account in place, with all their photos, videos, and conversation history. It works the same way with cloud email providers like Gmail or Microsoft Office 365, and with cloud storage providers like Dropbox or Google Drive.

For businesses, switching to cloud computing removes some IT costs and overhead: for instance, they no longer need to update and maintain their own servers, as the cloud vendor they are using will do that. This especially makes an impact for small businesses that may not have been able to afford their own internal infrastructure but can outsource their infrastructure needs affordably via the cloud. The cloud can also make it easier for companies to operate internationally, because employees and customers can access the same files and applications from any location.



The Benefits/Advantages of Using Cloud Computing:

- It's flexible
Due to the architecture of cloud computing, enterprises and their users can access cloud services from anywhere with an internet connection, scaling services up or down as needed.
- It's efficient
Enterprises can develop new applications and rapidly get them into production—without worrying about the underlying infrastructure.
- It offers strategic value
Because cloud providers stay on top of the latest innovations and offer them as services to customers, enterprises can get more competitive advantages—and a higher return on investment—than if they'd invested in soon-to-be obsolete technologies.
- It's secure
Enterprises often ask, What are the security risks of cloud computing? They are considered relatively low. Cloud computing security is generally recognized as stronger than that in enterprise data centers, because of the depth and breadth of the security mechanisms cloud providers put into place. Plus, cloud providers' security teams are known as top experts in the field
- It's cost-effective
Whatever cloud computing service model is used, enterprises only pay for the computing resources they use. They don't need to overbuild data center capacity to handle unexpected spikes in demand or business growth, and they can deploy IT staff to work on more strategic initiatives.
- It's more cost-effective than on-site server installations and can provide faster service than a traditional installation.
- In addition, cloud computing models are easy to scale. The cloud extends alongside an organization's fluctuating needs, making it easy to accommodate real-time changes in computing power needs.
- The cloud is also very reliable, offering dependable data backup for disaster recovery and business continuity.
- Businesses that leverage cloud environments do not need to maintain complex hardware, nor do they need to build solutions from scratch. The cloud allows teams to get projects up and running as soon as they receive executive sign off.



Cloud Computing Service Models

Cloud computing services come mainly in three types of service models:

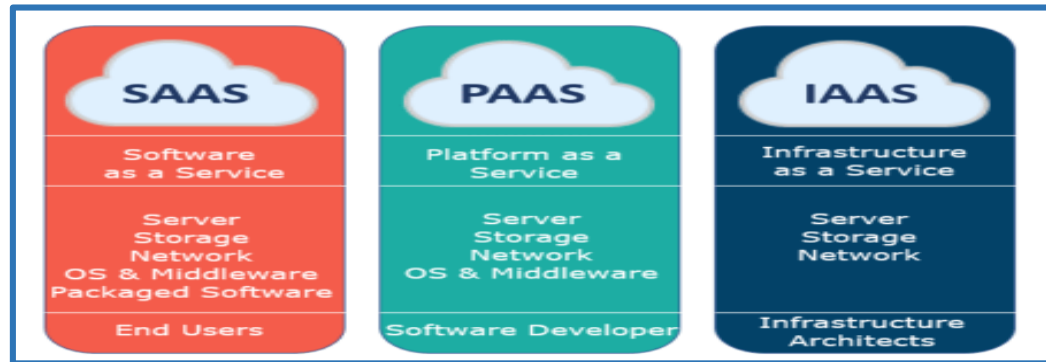
SaaS (Software as a Service),

IaaS (Infrastructure as a Service),

and PaaS (Platform as a Service).

Each of the cloud models has its own set of benefits that could serve the needs of various businesses.

Cloud service pricing models are categorized into pay per use, subscription-based and hybrid, which is a combination of pay-per-use and subscription pricing models.



Software as a Service

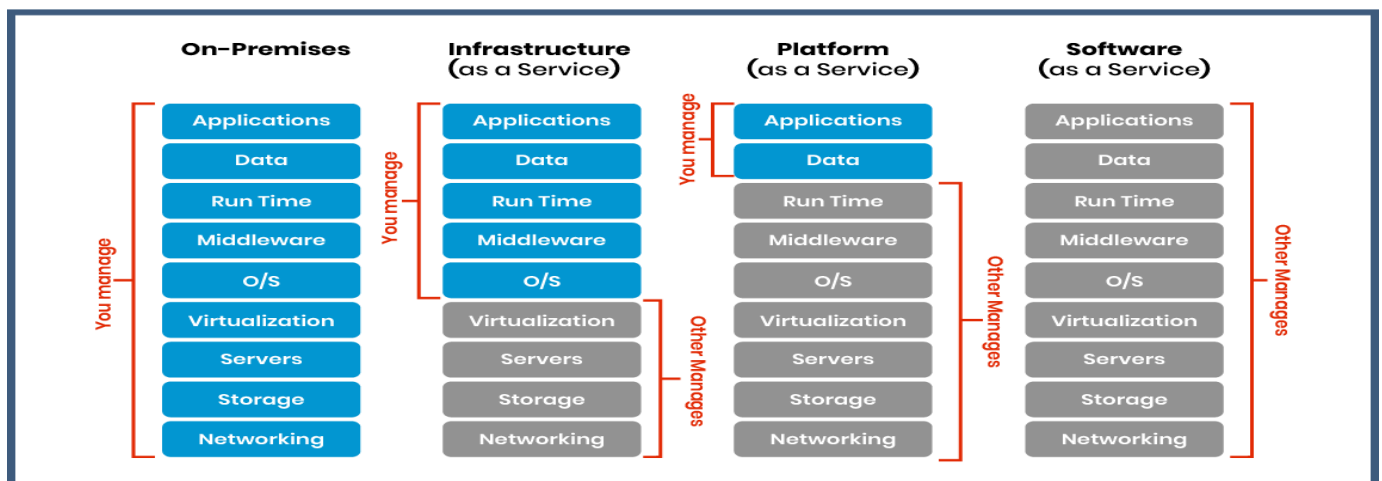
Software as a service vendors host the applications, making them available to users via the internet. With SaaS, businesses don't have to install or download any software to their existing IT infrastructures. SaaS ensures that users are always running the most up-to-date versions of the software. The SaaS provider handles maintenance and support.

Platform as a Service

Platform as a service offers developers a platform for software development and deployment over the internet, enabling them to access up-to-date tools. PaaS delivers a framework that developers can use to create customized applications. The organization or the PaaS cloud vendor manage the servers, storage and networking, while the developers manage the applications.

Infrastructure as a Service

Infrastructure as a service is used by companies that don't want to maintain their own on-premises data centers. IaaS provides virtual computing resources over the Internet. The IaaS cloud vendor hosts the infrastructure components that typically exist in an on-premises data center, including servers, storage and networking hardware, as well as the hypervisor or virtualization layer.



Pros of using SaaS include:

- **Easy to access and use:** The main benefit of SaaS products is that organizations can use them as soon as they subscribe because it's the easiest cloud model to set up and run. SaaS is also the easiest to maintain because the cloud providers manage everything.
- **Scalability:** To add users, organizations just have to upgrade their existing plans or subscriptions. They don't have to buy additional server space or software licenses.

Cons of using SaaS include:

- **Lack of control:** Organizations don't have control over their providers' cloud infrastructures. Consequently, if the provider has an outage, they do as well.
- **Issue with integrations:** Organizations may have trouble integrating their existing in-house software with the SaaS applications, as their in-house APIs and data structures may not integrate with the external applications.

Pros of PaaS include:

- **Simplicity, convenience:** PaaS providers deliver most of the infrastructure and other IT services for organizations, which users can access as long as they have an internet connection and a web browser.
- **Faster development:** PaaS platforms provide compute and storage infrastructures, along with text editing, version management, compiling and testing services that help developers efficiently build new software. They also help development teams work together, regardless of wherever they are physically located.
- **Cons of PaaS include:**
- **Lack of scalability:** PaaS tools are a little more rigid than IaaS tools, which may be an issue for organizations that experience extremely high demand for their products or services at various times throughout the year.
- **Vendor lock-in:** Because PaaS vendors have unique configuration requirements, organizations may find it difficult to move from one provider to another.

What is AWS and its Benefits?

Amazon web service is an online platform that provides scalable and cost-effective cloud computing solutions. AWS (Amazon Web Services) is a comprehensive, evolving cloud computing platform provided by Amazon that includes a mixture of infrastructure as a service (IaaS), platform as a service (PaaS) and packaged software as a service (SaaS) offerings. AWS services can offer an organization tools such as compute power, database storage and content delivery services.

AWS launched in 2006 from the internal infrastructure that Amazon.com built to handle its online retail operations. AWS was one of the first companies to introduce a pay-as-you-go cloud computing model that scales to provide users with compute, storage or throughput as needed.

AWS offers many different tools and solutions for enterprises and software developers that can be used in data centers in up to 190 countries. Groups such as government agencies, education institutions, nonprofits and private organizations can use AWS services.

Advantages of AWS

Following are the pros of using AWS services:

- AWS allows organizations to use the already familiar programming models, operating systems, databases, and architectures.
- It is a cost-effective service that allows you to pay only for what you use, without any up-front or long-term commitments.
- You will not require to spend money on running and maintaining data centers.
- Offers fast deployments
- You can easily add or remove capacity.
- You are allowed cloud access quickly with limitless capacity.
- Total Cost of Ownership is very low compared to any private/dedicated servers.
- Offers Centralized Billing and management
- Offers Hybrid Capabilities
- Allows you to deploy your application in multiple regions around the world with just a few clicks

Disadvantages of AWS

- If you need more immediate or intensive assistance, you'll have to opt for paid support packages.
- Amazon Web Services may have some common cloud computing issues when you move to a cloud. For example, downtime, limited control, and backup protection.
- AWS sets default limits on resources which differ from region to region. These resources consist of images, volumes, and snapshots.
- Hardware-level changes happen to your application which may not offer the best performance and usage of your applications.

Applications of AWS services

Amazon Web services are widely used for various computing purposes like:

- Web site hosting
- Application hosting/SaaS hosting
- Media Sharing (Image/ Video)
- Mobile and Social Applications
- Content delivery and Media Distribution
- Storage, backup, and disaster recovery
- Development and test environments
- Academic Computing
- Search Engines
- Social Networking


AWS Global Infrastructure

The AWS global infrastructure is massive and is divided into geographical regions. The geographical regions are then divided into separate availability zones. While selecting the geographical regions for AWS, three factors come into play

- Optimizing Latency
- Reducing cost
- Government regulations (Some services are not available for some regions)

AWS Regions

- An **AWS Region** is a geographical area.
 - **Data replication** across Regions is controlled by you.
 - **Communication** between Regions uses AWS backbone network infrastructure.
- Each Region provides full redundancy and connectivity to the network.
- A Region typically consists of two or more **Availability Zones**.



Example: London Region

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

- The AWS Cloud infrastructure is built around Regions. AWS has 22 Regions worldwide. An **AWS Region** is a physical geographical location with one or more **Availability Zones**. Availability Zones in turn consist of one or more **data centers**.
- To achieve fault tolerance and stability, Regions are isolated from one another. Resources in one Region are not automatically replicated to other Regions. When you store data in a specific Region, it is not replicated outside that Region.
- It is your responsibility to replicate data across Regions, if your business needs require it.
- AWS Regions that were introduced before March 20, 2019 are enabled by default. Regions that were introduced after March 20, 2019—such as Asia Pacific (Hong Kong) and Middle East (Bahrain)—are disabled by default. You must enable these Regions before you can use them. You can use the AWS Management Console to enable or disable a Region.
- Some Regions have restricted access. An Amazon AWS (**China**) account provides access to the Beijing and Ningxia Regions only. To learn more about AWS in China, see: <https://www.amazonaws.cn/en/about-aws/china/>. The isolated **AWS GovCloud (US)** Region is designed to allow US government agencies and customers to move sensitive workloads into the cloud by addressing their specific regulatory and compliance requirements.
- **For accessibility:** Snapshot from the infrastructure.aws website that shows a picture of downtown London including the Tower Bridge and the Shard. It notes that there are three Availability Zones in the London region. **End of accessibility description.**

Selecting a Region



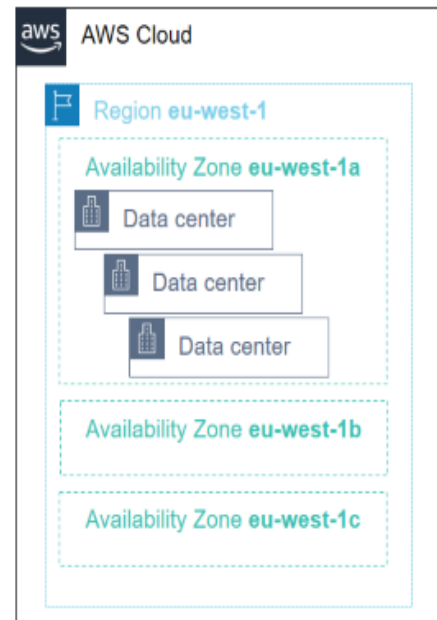
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

8

- There are a few factors that you should consider when you select the optimal Region or Regions where you store data and use AWS services.
- One essential consideration is **data governance and legal requirements**. Local laws might require that certain information be kept within geographical boundaries. Such laws might restrict the Regions where you can offer content or services. For example, consider the European Union (EU) Data Protection Directive.
- All else being equal, it is generally desirable to run your applications and store your data in a Region that is as close as possible to the user and systems that will access them. This will help you **reduce latency**. CloudPing is one website that you can use to test latency between your location and all AWS Regions. To learn more about CloudPing, see: <http://www.cloudping.info/>
- Finally, there is some variation in the cost of running services, which can depend on which Region you choose. For example, as of this writing, running an On-Demand t3.medium size Amazon Elastic Compute Cloud (Amazon EC2) Linux instance in the US East (Ohio) Region costs \$0.0416 per hour, but running the same instance in the Asia Pacific (Tokyo) Region costs \$0.0544 per hour.

Availability Zones

- Each **Region** has multiple Availability Zones.
- Each **Availability Zone** is a fully isolated partition of the AWS infrastructure.
 - Availability Zones consist of discrete **data centers**
 - They are designed for fault isolation
 - They are interconnected with other Availability Zones by using high-speed private networking
 - You choose your Availability Zones.
 - **AWS recommends replicating data and resources across Availability Zones** for resiliency.



- Each AWS Region has multiple, isolated locations that are known as Availability Zones.
- Each Availability Zone provides the ability to operate applications and databases that are more highly available, fault-tolerant, and scalable than would be possible with a single data center. Each Availability Zone can include multiple data centers (typically three), and at full-scale, they can include hundreds of thousands of servers. They are fully isolated partitions of the AWS Global Infrastructure. Availability Zones have their own power infrastructure, and they are physically separated by many kilometers from other Availability Zones—though all Availability Zones are within 100 km of each other.
- All Availability Zones are interconnected with high-bandwidth, low-latency networking over fully redundant, dedicated fiber that provides high-throughput between Availability Zones. The network accomplishes synchronous replication between Availability Zones.
- Availability Zones help build highly available applications. When an application is partitioned across Availability Zones, companies are better isolated and protected from issues such as lightning, tornadoes, earthquakes, and more.
- You are responsible for selecting the Availability Zones where your systems will reside. Systems can span multiple Availability Zones. AWS recommends replicating across Availability Zones for resiliency. You should design your systems to survive the temporary or prolonged failure of an Availability Zone if a disaster occurs.

AWS data centers

- AWS data centers are **designed for security**.
- Data centers are where the data resides and data processing occurs.
- Each data center has redundant power, networking, and connectivity, and is housed in a separate facility.
- A data center typically has 50,000 to 80,000 physical servers.



The foundation for the AWS infrastructure is the data centers. Customers do not specify a data center for the deployment of resources. Instead, an Availability Zone is the most granular level of specification that a customer can make. However, a data center is the location where the actual data resides. Amazon operates state-of-the-art, highly available data centers. Although rare, failures can occur that affect the availability of instances in the same location. If you host all your instances in a single location that is affected by such a failure, none of your instances will be available.

Data centers are securely designed with several factors in mind:

Each location is carefully evaluated to **mitigate environmental risk**.

- Data centers have a **redundant design** that anticipates and tolerates failure while maintaining service levels.
- To ensure availability, **critical system components are backed up** across multiple Availability Zones.
- To ensure capacity, AWS continuously monitors service usage to deploy infrastructure to support availability commitments and requirements.
- Data center **locations are not disclosed** and all access to them is restricted.
- In case of failure, automated processes move data traffic away from the affected area.

AWS uses **custom network equipment** sourced from **multiple original device manufacturers (ODMs)**. ODMs design and manufacture products based on specifications from a second company. The second company then rebrands the products for sale.

Points of Presence

- AWS provides a global network of **Points of Presence** locations
- Consists of **edge locations** and a much smaller number of **Regional edge caches**
- Used with Amazon CloudFront
 - A global Content Delivery Network (CDN), that delivers content to end users with **reduced latency**
- Regional edge caches used for content with infrequent access.



aws

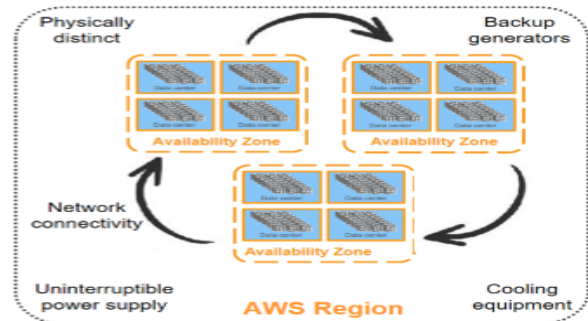
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

11

- **Amazon CloudFront** is a **content delivery network (CDN)** used to distribute content to end users to reduce latency. **Amazon Route 53** is a Domain Name System (DNS) service. Requests going to either one of these services will be routed to the nearest **edge location** automatically in order to lower latency.
- **AWS Points of Presence** are located in most of the major cities around the world. By **continuously measuring internet connectivity, performance and computing to find the best way to route requests**, the Points of Presence deliver a better near real-time user experience. They are used by many AWS services, including Amazon CloudFront, Amazon Route 53, AWS Shield, and AWS Web Application Firewall (AWS WAF) services.
- **Regional edge caches** are used by default with Amazon CloudFront. Regional edge caches are used when you have content that is not accessed frequently enough to remain in an **edge location**. Regional edge caches absorb this content and provide an alternative to that content having to be fetched from the origin server.

AWS infrastructure features

- **Elasticity and scalability**
 - Elastic infrastructure; dynamic adaption of capacity
 - Scalable infrastructure; adapts to accommodate growth
- **Fault-tolerance**
 - Continues operating properly in the presence of a failure
 - Built-in redundancy of components
- **High availability**
 - High level of operational performance
 - Minimized downtime
 - No human intervention



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

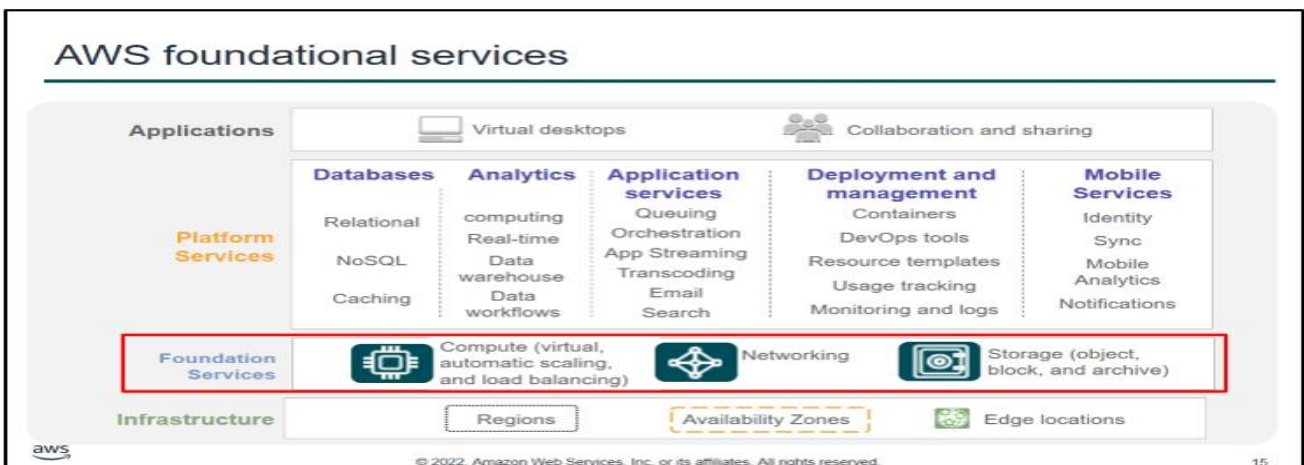
12

The AWS Global Infrastructure has several valuable features:

- First, it is **elastic** and **scalable**. This means resources can dynamically adjust to increases or decreases in capacity requirements. It can also rapidly adjust to accommodate growth.
- Second, this infrastructure is **fault tolerant**, which means it has built-in component redundancy which enables it to continue operations despite a failed component.
- Finally, it requires minimal to no human intervention, while providing **high availability** with minimal down time.

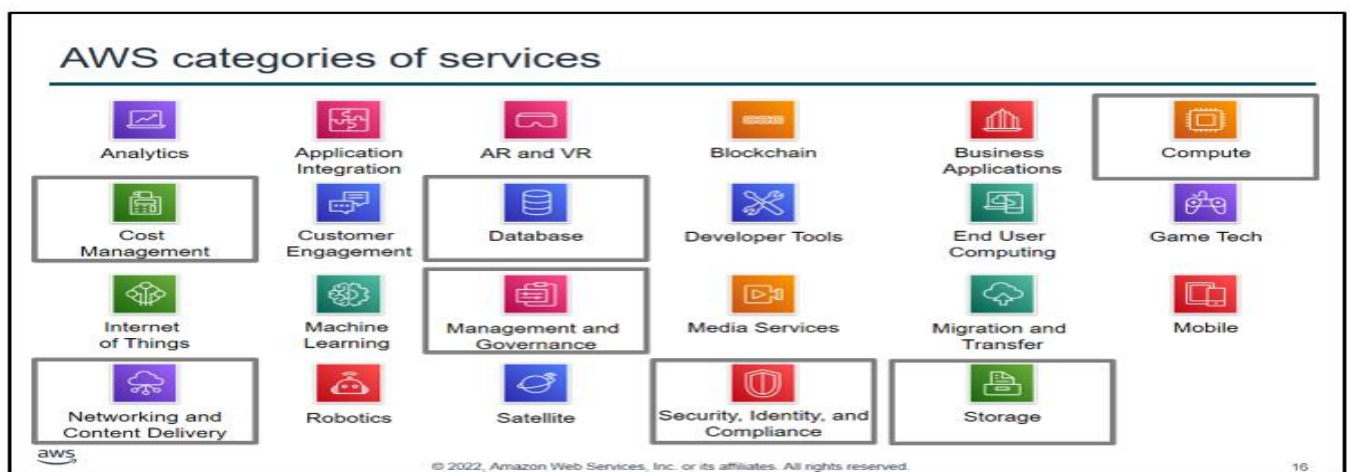
AWS Service and Service Category Overview.

AWS offers a broad set of global cloud-based products that can be used as building blocks for common cloud architectures. Here is a look at how these cloud based products are organized.



AWS Global Infrastructure can be broken down into three elements: Regions, Availability Zones, and Points of Presence, which include edge locations. This infrastructure provides the platform for a broad set of services, such as networking, storage, compute services, and databases—and these services are delivered as an on-demand utility that is available in seconds, with pay-as-you-go pricing.

For accessibility: Marketing diagram showing infrastructure at the bottom, consisting of Regions, Availability Zones, and edge locations. The next level up is labeled Foundational Services and includes graphics for compute, networking, and storage. That level is highlighted. Next level up is platform services that includes databases, analytics, app services, deployment and management, and mobile services. Top layer is labeled applications and includes virtual desktops and collaboration and sharing. **End of accessibility description.**



AWS offers a broad set of cloud-based services. There are 23 different product or service categories, and each category consists of one or more services. This course will not attempt to introduce you to each service. Rather, the focus of this course is on the services that are most widely used and offer the best introduction to the AWS Cloud. This course also focuses on services that are more likely to be covered in the AWS Certified Cloud Practitioner exam.

The categories that this course will discuss are highlighted on the slide: Compute, Cost Management, Database, Management and Governance, Networking and Content Delivery, Security, Identity, and Compliance, and Storage.

Storage service category



Photo from: <https://www.pexels.com/photo/black-and-grey-device-1592892/>



AWS storage services



Amazon Simple
Storage Service
(Amazon S3)



Amazon Elastic
Block Store
(Amazon EBS)



Amazon Elastic
File System
(Amazon EFS)



Amazon Simple
Storage Service
Glacier



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

17

AWS storage services include the services listed here, and many others.

- **Amazon Simple Storage Service (Amazon S3)** is an object storage service that offers scalability, data availability, security, and performance. Use it to store and protect any amount of data for websites, mobile apps, backup and restore, archive, enterprise applications, Internet of Things (IoT) devices, and big data analytics.
- **Amazon Elastic Block Store (Amazon EBS)** is high-performance block storage that is designed for use with Amazon EC2 for both throughput and transaction intensive workloads. It is used for a broad range of workloads, such as relational and non-relational databases, enterprise applications, containerized applications, big data analytics engines, file systems, and media workflows.
- **Amazon Elastic File System (Amazon EFS)** provides a scalable, fully managed elastic Network File System (NFS) file system for use with AWS Cloud services and on-premises resources. It is built to scale on demand to petabytes, growing and shrinking automatically as you add and remove files. It reduces the need to provision and manage capacity to accommodate growth.
- **Amazon Simple Storage Service Glacier** is a secure, durable, and extremely low-cost Amazon S3 cloud storage class for data archiving and long-term backup. It is designed to deliver 11 9s of durability, and to provide comprehensive security and compliance capabilities to meet stringent regulatory requirements.

Compute service category



Photo from <https://www.pexels.com/photo/technology-computer-lines-board-50711/>



AWS Compute services



Amazon EC2



Amazon
EC2
Auto Scaling



Amazon Elastic
Container Service
(Amazon ECS)



Amazon EC2
Container
Registry



AWS Elastic
Beanstalk



AWS
Lambda



Amazon Elastic
Kubernetes Service
(Amazon EKS)



AWS
Fargate



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

18

AWS compute services include the services listed here, and many others.

- **Amazon Elastic Compute Cloud (Amazon EC2)** provides resizable compute capacity as virtual machines in the cloud.
- **Amazon EC2 Auto Scaling** enables you to automatically add or remove EC2 instances according to conditions that you define.
- **Amazon Elastic Container Service (Amazon ECS)** is a highly scalable, high-performance container orchestration service that supports Docker containers.
- **Amazon Elastic Container Registry (Amazon ECR)** is a fully-managed Docker container registry that makes it easy for developers to store, manage, and deploy Docker container images.
- **AWS Elastic Beanstalk** is a service for deploying and scaling web applications and services on familiar servers such as Apache and Microsoft Internet Information Services (IIS).
- **AWS Lambda** enables you to run code without provisioning or managing servers. You pay only for the compute time that you consume. There is no charge when your code is not running.
- **Amazon Elastic Kubernetes Service (Amazon EKS)** makes it easy to deploy, manage, and scale containerized applications that use Kubernetes on AWS.
- **AWS Fargate** is a compute engine for Amazon ECS that allows you to run containers without having to manage servers or clusters.

Database service category



Photo from <https://aws.amazon.com/compliance/data-center/data-centers/>



AWS Database services



Amazon Relational
Database Service



Amazon Aurora



Amazon
Redshift



Amazon
DynamoDB



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

19

AWS database services include the services listed here, and many others.

- **Amazon Relational Database Service (Amazon RDS)** makes it easy to set up, operate, and scale a relational database in the cloud. It provides resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups.
- **Amazon Aurora** is a MySQL and PostgreSQL-compatible relational database. It is up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases.
- **Amazon Redshift** enables you to run analytic queries against petabytes of data that is stored locally in Amazon Redshift, and directly against exabytes of data that are stored in Amazon S3. It delivers fast performance at any scale.
- **Amazon DynamoDB** is a key-value and document database that delivers single-digit millisecond performance at any scale, with built-in security, backup and restore, and in-memory caching.

Networking and content delivery service category



Photo by Umberto on Unsplash



**AWS networking
and content delivery services**



Amazon VPC



Elastic Load
Balancing



Amazon
CloudFront



AWS Transit
Gateway



Amazon
Route 53



AWS Direct
Connect



AWS VPN



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

20

AWS networking and content delivery services include the services listed here, and many others.

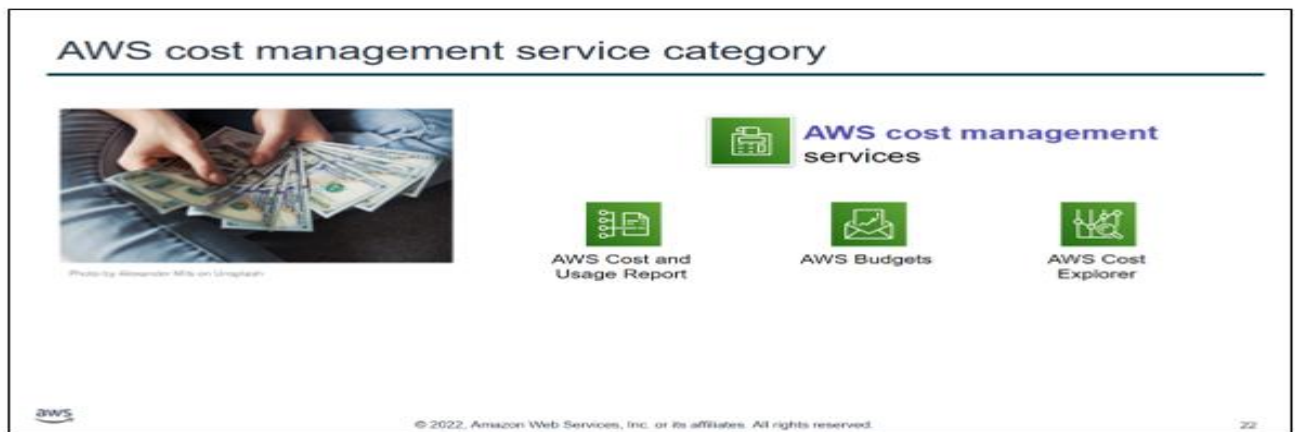
- **Amazon Virtual Private Cloud (Amazon VPC)** enables you to provision logically isolated sections of the AWS Cloud.
- **Elastic Load Balancing** automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions.
- **Amazon CloudFront** is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and application programming interfaces (APIs) to customers globally, with low latency and high transfer speeds.
- **AWS Transit Gateway** is a service that enables customers to connect their Amazon Virtual Private Clouds (VPCs) and their on-premises networks to a single gateway.
- **Amazon Route 53** is a scalable cloud Domain Name System (DNS) web service designed to give you a reliable way to route end users to internet applications. It translates names (like `www.example.com`) into the numeric IP addresses (like `192.0.2.1`) that computers use to connect to each other.
- **AWS Direct Connect** provides a way to establish a dedicated private network connection from your data center or office to AWS, which can reduce network costs and increase bandwidth throughput.
- **AWS VPN** provides a secure private tunnel from your network or device to the AWS global network.

Security, identity, and compliance service category



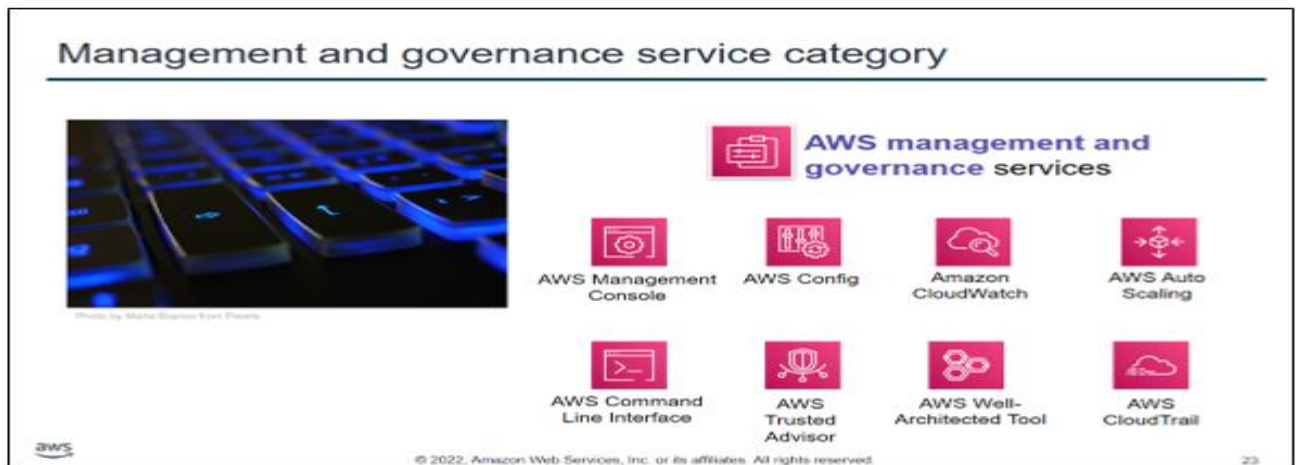
AWS security, identity, and compliance services include the services listed here, and many others.

- **AWS Identity and Access Management (IAM)** enables you to manage access to AWS services and resources securely. By using IAM, you can create and manage AWS users and groups. You can use IAM permissions to allow and deny user and group access to AWS resources.
- **AWS Organizations** allows you to restrict what services and actions are allowed in your accounts.
- **Amazon Cognito** lets you add user sign-up, sign-in, and access control to your web and mobile apps.
- **AWS Artifact** provides on-demand access to AWS security and compliance reports and select online agreements.
- **AWS Key Management Service (AWS KMS)** enables you to create and manage keys. You can use AWS KMS to control the use of encryption across a wide range of AWS services and in your applications.
- **AWS Shield** is a managed Distributed Denial of Service (DDoS) protection service that safeguards applications running on AWS.



AWS cost management services include the services listed here, and others.

- The **AWS Cost and Usage Report** contains the most comprehensive set of AWS cost and usage data available, including additional metadata about AWS services, pricing, and reservations.
- **AWS Budgets** enables you to set custom budgets that alert you when your costs or usage exceed (or are forecasted to exceed) your budgeted amount.
- **AWS Cost Explorer** has an easy-to-use interface that enables you to visualize, understand, and manage your AWS costs and usage over time.



AWS management and governance services include the services listed here, and others.

- The **AWS Management Console** provides a web-based user interface for accessing your AWS account.
- **AWS Config** provides a service that helps you track resource inventory and changes.
- **Amazon CloudWatch** allows you to monitor resources and applications.
- **AWS Auto Scaling** provides features that allow you to scale multiple resources to meet demand.
- **AWS Command Line Interface** provides a unified tool to manage AWS services.
- **AWS Trusted Advisor** helps you optimize performance and security.
- **AWS Well-Architected Tool** provides help in reviewing and improving your workloads.
- **AWS CloudTrail** tracks user activity and API usage.