# Module 1: Introduction to AWS

A client can be a web browser or desktop application that a person interacts with to make requests to computer servers. A server can be services such as Amazon Elastic Compute Cloud (Amazon EC2), a type of virtual server.

## Cloud computing

Cloud computing is the on-demand delivery of IT resources over the internet with pay-as-you-go pricing. The undifferentiated heavy lifting of IT are all complex IT activities that don't separate you from your competition, such as installing a MySQL server.

**Burst capacity:** Use when on premise infrastructure takes heavy loads on certain times.

### Deployment models for cloud computing

1. **Cloud-based:** Runs all parts of the app in the cloud.
2. **On-premises:** Also known as a private cloud deployment. In this model, resources are deployed on premises by using virtualization and resource management tools.
3. **Hybrid:** Cloud-based resources are connected to on-premises infrastructure.

### Benefits of cloud computing

1. **Trade upfront expense for variable expense:** Variable expense means you only pay for computing resources you consume instead of investing in resources you may not need.
2. **Stop spending money to run and maintain data centers**
3. **Stop guessing capacity:** Instead of paying for unused resources or having to deal with limited capacity, you can access only the capacity that you need. You can also scale in or scale out in response to demand.
4. **Benefit from massive economies of scale**
5. **Increase speed and agility:** The flexibility of cloud computing makes it easier for you to develop and deploy applications.
6. **Go global in minutes:** The global footprint of the AWS Cloud enables you to deploy applications to customers around the world quickly, while providing them with low latency.

Concepts in cloud computing

**Resiliency:** refers to the ability of a solution to absorb the impact of a problem in one or more parts of a system, while continuing to provide an acceptable service level to the business.

**Elasticity:** Ability to scale resources.

**Flexibility:** Ability to use a broad range of products

**Agility:** Ability to release products faster.

# Module 2: Compute in the cloud

**Multitenancy:** Sharing underlying hardware between virtual machines.

**Vertical scaling:** Give an instance more memory or more CPU depending on needs.

## Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Compute Cloud (Amazon EC2) provides secure, resizable compute capacity in the cloud as Amazon EC2 instances. EC2s can be copied to other regions. Characteristics:

* Allows the customer to retain full administrative privileges of the underlying virtual infrastructure
* You pay only for the compute time you use when an instance is running, not when it is stopped or terminated.

### How Amazon EC2 works

1. Launch: You launch an instance by selecting a template with basic configurations. These configurations include:

* Operating system
* Application server or applications
* Instance type, which is the specific hardware configuration of your instance.
* Security settings to control the network traffic that can flow into and out of your instance.

1. Connect: Your programs and applications have multiple different methods to connect directly to the instance and exchange data.
2. Use

### Instance types

* **General purpose:** Provides a balance of compute, memory, and networking resources.
* **Compute optimized:** Ideal for compute-bound applications that benefit from high-performance processors.
* **Memory optimized:** Designed to deliver fast performance for workloads that process large datasets in memory.
* **Accelerated computing:** Uses hardware accelerators, or coprocessors, to perform some functions more efficiently than is possible in software running on CPUs. Ideal for workloads such as graphics applications, game streaming, and application streaming.
* **Storage optimized:** Designed for workloads that require high, sequential read and write access to large datasets on local storage. Ideal for data warehousing applications.

The term input/output operations per second (**IOPS**) is a metric that measures the performance of a storage device.

### Pricing

* **On-Demand:** Ideal for short-term, irregular workloads that cannot be interrupted. No upfront costs or minimum contracts apply. The instances run continuously until you stop them, and you pay for only the compute time you use. Instance type, AMI type and Region determine cost.
* **Amazon EC2 Savings Plans:** Reduce your compute costs by committing to a consistent amount of compute usage for a 1-year or 3-year term.
* **Reserved Instances:** Billing discount applied to the use of On-Demand Instances in your account. Extend boundaries across different Availability Zones.
* **Spot Instances:** Ideal for workloads with flexible start and end times, or that can withstand interruptions. Spot Instances use unused Amazon EC2 computing capacity. Useful for dev/test environments. If the instance is stopped by Amazon in the first hour, this first hour is not billed.
* **Dedicated Hosts:** Physical servers with Amazon EC2 instance capacity that is fully dedicated to your use.

### Scalability

Scalability involves beginning with only the resources you need and designing your architecture to automatically respond to changing demand by scaling out or in.

### EC2 Auto Scaling

Enables you to automatically add or remove Amazon EC2 instances in response to your applications demand. By automatically scaling your instances in and out as needed, you are able to maintain a greater sense of application availability.

Within Amazon EC2 Auto Scaling, you can use two approaches:

1. Dynamic scaling responds to changing demand.
2. Predictive scaling automatically schedules the right number of EC2 instances based on predicted demand.

When you create an Auto Scaling group, you can set the **minimum** number of Amazon EC2 instances which is the number of instances that launch immediately after you have created the Auto Scaling group. Next, you can set the **desired** capacity, and then the **maximum** capacity.

### Elastic Load Balancing

Elastic Load Balancing is the AWS service that automatically distributes incoming application traffic across multiple resources.

A load balancer acts as a single point of contact for all incoming web traffic to your Auto Scaling group. This means that as you add or remove EC2 instances in response to the amount of incoming traffic, these requests route to the load balancer first. Load Balancers now support invoking Lambda functions to serve HTTP(S) requests.

**Application Load Balancer:** Apt for http and https traffic load balancing.

### Monolithic applications and microservices

Monolithic application: Application with tightly coupled components. In this approach to application architecture, if a single component fails, other components fail, and possibly the entire application fails.

Microservices approach: Application components are loosely coupled. If a single component fails, the other components continue to work because they are communicating with each other.

#### Amazon Simple Notification Service (Amazon SNS)

Is a publish/subscribe service. Using Amazon SNS topics, a publisher publishes messages to subscribers. SNS messages can be sent to registered addresses through email and lambdas.

#### Amazon Simple Queue Service (Amazon SQS)

Is a message queuing service. Using Amazon SQS, you can send, store, and receive messages between software components, without losing messages or requiring other services to be available. In Amazon SQS, an application sends messages into a queue. A user or service retrieves a message from the queue, processes it, and then deletes it from the queue. Implements messaging that is a typical integration pattern to **decouple application components**.

### Serverless computing

Your code runs on servers, but you do not need to provision, manage or manage these servers. Another benefit is the flexibility to scale serverless applications automatically. Serverless computing can adjust the applications' capacity by modifying the units of consumptions, such as throughput and memory.

#### AWS Lambda

Service that lets you run code without needing to provision or manage servers. While using AWS Lambda, you pay only for the compute time that you consume.

**AWS Lambda@Edge** is a serverless service that makes it possible to run event-triggered functions on Edge Locations within the AWS Content Delivery Network. Using AWS CloudFront, an administrator can introduce decision-making and compute processing closer to the viewer’s location.

### Containers

Standard way to package an application's code and dependencies into a single object. You can also use containers for processes in which there are essential requirements for security, reliability, and scalability. Container orchestration services help you to deploy, manage, and scale your containerized applications.

#### Amazon Elastic Container Service (Amazon ECS)

Highly scalable, high-performance container management system that enables you to run and scale containerized applications on AWS.

Amazon ECS supports Docker containers. Docker is a software platform that enables you to build, test, and deploy applications quickly. With Amazon ECS, you can use API calls to launch and stop Docker-enabled applications.

#### Amazon Elastic Kubernetes Service (Amazon EKS)

Fully managed service that you can use to run Kubernetes on AWS.

Kubernetes is open-source software that enables you to deploy and manage containerized applications at scale.

#### AWS Fargate

Serverless compute engine for containers. It works with both Amazon ECS and Amazon EKS. Manages your server infrastructure for you. You pay only for the resources that are required to run your containers.

### Other services

**Code Commit:** Helps in hosting git-based repositories securely.

**Code Deploy:** is a deployment service that allows developers to automate the installation of applications. AWS CodeDeploy can enable the update of those applications.

**Code Guru:** Development tool powered by ML, that provides code quality improvement recommendations.

**Code Pipeline:** Management tool that facilitates automation of various phases of the release process. Does not provide infrastructure.

**Code Star:** Unified user interface. Helps in swift and quick development, build and deployment of applications in AWS.

**Kinesis:** Makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information.

**Kinesis Data Analytics:** A fully managed solution and you want to use SQL to process the data from your data stream.

**Lightsail:** is an easy-to-use cloud platform that offers you everything needed to build an application or website, plus a cost-effective, monthly plan. “Upgrade to EC2” is the feature that allows customers to “create copy in EC2”.

**ParallelCluster:** is an open-source cluster management tool supported by AWS that helps customers deploy and manage HPC clusters in AWS.

**Quick Starts:** Helps in automating deployments aligned with the best practices. CloudFormation templates are included.

**Transfer Family:** SFTP, FTPS, FTP.

**X-Ray:** Helps developers analyze and debug production, distributed applications, such as those built using a microservices architecture. Can be used to detect performance issues for AWS Lambda applications.

# Module 3: Global infrastructure and reliability

## Selecting a Region

* Compliance with data governance and legal requirements: Depending on your company and location, you might need to run your data out of specific areas.
* Proximity to your customers: Selecting a Region that is close to your customers will help you to get content to them faster.
* Available services within a Region: The closest Region might not have all the features that you want to offer to customers.
* Pricing: May change based on local taxes or other factors.

## Availability zone

Single data center or a **group** of data centers within a Region. Availability Zones are located tens of miles apart from each other. This is close enough to have low latency (the time between when content requested and received) between Availability Zones. Regions are completely isolated.

## Edge locations

Site that **Amazon CloudFront** (a global content delivery service, that cache dynamic content and cant use S3 buckets as their origin server) uses to store cached copies of your content closer to your customers for faster delivery.

## Ways to interact with AWS services

**AWS Outposts:** Extend AWS infrastructure and services to your on-premises data center. Service that provides hybrid experience by facilitating running AWS services and infrastructure on premises.

### AWS management console

Web-based interface for accessing and managing AWS services. The console includes wizards and automated workflows that can simplify the process of completing tasks. A web-based user interface that helps users to access and manage aspects of AWS services.

You can also use the AWS Console mobile application to perform tasks such as monitoring resources, viewing alarms, and accessing billing information. Multiple identities can stay logged into the AWS Console mobile app at the same time.

### AWS command line interface

Enables you to control multiple AWS services directly from the command line within one tool. AWS CLI is available for users on Windows, macOS, and Linux.

By using AWS CLI, you can automate the actions that your services and applications perform through scripts. Also, SSH allows users to interact with AWS services through AWS CLI.

### Software development kits

SDKs make it easier for you to use AWS services through an API designed for your programming language or platform. SDKs enable you to use AWS services with your existing applications or create entirely new applications that will run on AWS.

### AWS Elastic Beanstalk

You provide code and configuration settings, and Elastic Beanstalk deploys the resources necessary to perform the following tasks: Adjust capacity, Load balancing, Automatic scaling, Application health monitoring. A service for deploying and scaling web applications and services. For deploying and scaling webapps.

### AWS CloudFormation

You can treat your infrastructure as code. This means that you can build an environment by writing lines of code. AWS CloudFormation provisions your resources in a safe, repeatable manner, enabling you to frequently build your infrastructure and applications without having to perform manual actions or write custom scripts.

## Other Services

**Global accelerator:** Improves performance and availability. It doesn’t include content caching capability and is for apps that are not HTTP.

**Resource access manager:** Allows to share resources between multiple AWS accounts.

**Service Catalog:** Ideal for organizing, distributing, and provisioning application stacks on AWS. Uses AWS IAM & AWS CloudFormation to create a portfolio of products.

# Module 4: Networking

## Amazon Virtual Private Cloud (Amazon VPC)

A free networking service that you can use to establish boundaries around your AWS resources.

Amazon VPC enables you to provision an isolated section of the AWS Cloud. In this isolated section, you can launch resources in a virtual network that you define. Within a VPC, you can organize your resources into subnets. A **subnet** is a section of a VPC that can contain resources such as Amazon EC2 instances.

**Internet gateway:** To allow public traffic from the internet to access your VPC. Is a connection between a VPC and the internet. Without an internet gateway, no one can access the resources within your VPC.

### Virtual private gateway

To access private resources in a VPC, you can use a virtual private gateway. Is the component that allows protected internet traffic to enter the VPC.

**Virtual private network (VPN):** Connection that encrypts (or protects) your internet traffic from all the other requests around it. Full hour is billed.

A **virtual private gateway** enables you to establish a **virtual private network** (VPN) connection between your VPC and a private network, such as an on-premises data center or internal corporate network. A virtual private gateway allows traffic into the VPC only if it is coming from an approved network.

**Network address translation (NAT) gateway:** To enable instances in a private subnet to connect to the internet or other AWS services, but prevent the internet from initiating a connection with those instances.

## AWS Direct Connect

AWS Direct Connect is a service that enables you to establish a dedicated private connection between your data center and a VPC.

The private connection that AWS Direct Connect provides helps you to reduce network costs and increase the amount of bandwidth that can travel through your network. AWS Direct Connect supports only the BGP (Border Gateway Protocol) routing protocol for this connectivity.

## Subnets

A section of a VPC in which you can group resources based on security or operational needs. Subnets can be public or private. In a VPC, subnets can communicate with each other.

**Public subnets** contain resources that need to be accessible by the public, such as an online store’s website.

**Private subnets** contain resources that should be accessible only through your private network, such as a database that contains customers’ personal information and order histories.

## Network traffic in a VPC

A packet is a unit of data sent over the internet or a network, which is sent as request to an application hosted in the AWS Cloud. It enters a VPC through an internet gateway. Before a packet can enter a subnet or exit one, it checks for permissions. These permissions indicate who sent the packet and how the packet is trying to communicate with the resources in a subnet.

The VPC component that checks packet permissions for subnets is a network access control list (ACL). VPC endpoint configured by AWS Private Link can have Amazon SES (Simple Email Service.)

## Network access control lists (NACLs)

An optional virtual firewall that controls inbound and outbound traffic at the **subnet level**. Each AWS account includes a default network ACL. When configuring your VPC, you can use your account’s default network ACL or create custom network ACLs.

By default, your account’s default network ACL **allows all inbound and outbound traffic**, but you can modify it. For custom network ACLs, all inbound and outbound traffic is denied until you add rules to specify which traffic to allow. Additionally, all network ACLs have an explicit deny rule. This rule ensures that if a packet doesn’t match any of the other rules on the list, the packet is denied.

### Stateless packet filtering

**Network ACLs perform stateless packet filtering**. They remember nothing and check packets that cross the subnet border each way: inbound and outbound. The VPC component that checks packet permissions for an Amazon EC2 instance is a security group.

#### Security groups

A security group is a virtual firewall that controls inbound and outbound traffic for an Amazon EC2 instance. By default, a security group denies all inbound traffic and allows all outbound traffic. You can add custom rules to configure which traffic to allow or deny. **Instance level**.

If you have multiple instances within a subnet, you can associate them with the same security group.

### Stateful packet filtering

**Security groups perform stateful packet filtering**. They remember previous decisions made for incoming packets.

Both network ACLs and security groups enable you to configure custom rules for the traffic in your VPC.

|  |  |
| --- | --- |
| Private subnet | Isolates databases containing customers personal information |
| Virtual private gateway | Create a VPN connection between VPC and the internal corporate network |
| Public subnet | Support the customer facing website |
| AWS Direct connect | Establish a dedicated connection between the on-premises data center and the VPC |
| Security groups | They are stateless and deny all inbound traffic by default. |

## Domain Name System (DNS)

Customers enter a web address into their browser, and they can access the website. This happens because of Domain Name System (DNS) resolution. DNS resolution involves a DNS server communicating with a web server.

## Amazon Route 53

Amazon Route 53 is a DNS web service. It gives developers and businesses a reliable way to route end users to internet applications hosted in AWS. Amazon Route 53 connects user requests to infrastructure running in AWS. It can route users to infrastructure outside of AWS.

Another feature of Route 53 is the ability to manage the DNS records for domain names. You can register new domain names directly in Route 53. You can also transfer DNS records for existing domain names managed by other domain registrars. This enables you to manage all your domain names within a single location.

It takes 24 to 48 hours to update resolvers around the world because those can only reflect their changes in their cache after the Time to Live has expired.

The geolocation policy allows for different resources to serve content based on the origin of the request.

* **Multivalue routing:** Returns multiple healthy IP addresses. Improves availability and load balancing.
* **Failover routing:** When 99.99% of uptime is required. Used in disaster recovery scenarios.
* **Weighted routing:** Route traffic in proportions.

**VPC peering connection** is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account. The VPCs can be in different regions (also known as an inter-region VPC peering connection). fault-Tolerance, disaster recovery and redundancy.

## Other services

**Elastic IP:** Static IP address that can be created and assigned to an AWS Account rather than to an Availability Zone.

**EventBridge:** is a serverless event bus that makes it easy to connect applications together using data from your own applications.

# Module 5: Storage & Database

## Instance stores

Block-level storage volumes behave like physical hard drives. An instance store provides temporary block-level storage for an Amazon EC2 instance. An instance store is disk storage that is physically attached to the host computer for an EC2 instance, and therefore has the same lifespan as the instance. When the instance is terminated, you lose any data in the instance store.

**Amazon Elastic Block Store (Amazon EBS)** is a service that provides block-level storage volumes. If you stop or terminate an Amazon EC2 instance, all the data on the attached EBS volume remains available.

To create an EBS volume, you define the configuration (such as volume size and type) and provision it. After you create an EBS volume. Because EBS volumes are for data that needs to persist, it’s important to back up the data. You can take incremental backups of EBS volumes by creating Amazon EBS snapshots. It is replicated in the same Availability Zone.

### Amazon EBS snapshots

An EBS snapshot is an incremental backup. This means that the first backup taken of a volume copies all the data. For subsequent backups, only the blocks of data that have changed since the most recent snapshot are saved. In full backups, all the data in a storage volume copies each time a backup occurs. Store snapshots on S3 can help reduce costs. Responsibility of the customer.

Data is not saved in backups outside a region unless you configure it that way. EBS snapshots are customer’s responsibility.

## Object storage

In object storage, each object consists of data, metadata, and a key.

* The data might be an image, video, text document, or any other type of file.
* Metadata contains information about what the data is, how it is used, the object size, and so on. Instance metadata is the defined parameters and attributes specified in instance configuration. User data is information that is passed to the instance’s operating system to automatically execute during boot time.
* An object’s key is its unique identifier.

## Amazon Simple Storage Service (Amazon S3)

A service that provides object-level storage. Amazon S3 stores data as objects in buckets. Amazon S3 offers unlimited storage space. The maximum file size for an object in Amazon S3 is 5 TB (Virtually unlimited storage).

When you upload a file to Amazon S3, you can set permissions to control visibility and access to it. You can also use the Amazon S3 versioning feature to track changes to your objects over time. Lifecycle transition requests (GET/PUT) are paid, but the number of buckets do not add cost.

By default, a user can only access resources they created. S3 supports both server-side and client-side encryptions.

### Amazon S3 storage classes

With Amazon S3, you pay only for what you use. You can choose from a range of storage classes to select a fit for your business and cost needs. When selecting an Amazon S3 storage class, consider these two factors: How often you plan to retrieve your data & How available you need your data to be.

* **S3 Standard:** Provides high availability for objects. Has a higher cost than other storage classes intended for infrequently accessed data and archival storage. Stores data in a minimum of three Availability Zones. Object versioning and lifecycle policies (obj transition to other storage class after time).
* **S3 Standard-Infrequent Access (S3 Standard-IA):** Ideal for infrequently accessed data. Similar to S3 Standard but has a lower storage price and higher retrieval price. Stores data in a minimum of three Availability Zones.
* **S3 One Zone-Infrequent Access (S3 One Zone-IA):** Stores data in a single Availability Zone
* **S3 Intelligent-Tiering:** Ideal for data with unknown or changing access patterns. Requires a small monthly monitoring and automation fee per object. If you haven’t accessed an object for 30 consecutive days, Amazon S3 automatically moves it to the infrequent access tier, S3 Standard-IA. If you access an object in the infrequent access tier, Amazon S3 automatically moves it to the frequent access tier, S3 Standard.
* **S3 Glacier:** Low-cost storage designed for data archiving. Able to retrieve objects within a few minutes to hours
* **S3 Glacier Deep Archive:** Lowest-cost object storage class ideal for archiving. Able to retrieve within 12 hours

## Athena

Is a serverless query service used to analyze BigData stored in S3. It queries data directly from Amazon S3 and there are no additional data storage commitments beyond the object storage. It is compatible with data formats such as CSV, JSON, ORC, AVRO and Parquet.

## File storage

In file storage, multiple clients can access data that is stored in shared file folders. In this approach, a storage server uses block storage with a local file system to organize files. Clients access data through file paths.

Compared to block storage and object storage, file storage is ideal for use cases in which a large number of services and resources need to access the same data at the same time.

**Amazon Elastic File System (Amazon EFS)** is a serverless scalable file system used with AWS Cloud services and on-premises resources. As you add and remove files, Amazon EFS grows and shrinks automatically. It can scale on demand without disrupting applications. Stores data **across multiple Availability Zone**. Up to petabytes. You can also share files between thousands of Amazon EC2 instances and on-premises servers via AWS Direct Connect or AWS VPN. Regional offering.

## Relational databases

In a relational database, data is stored in a way that relates it to other pieces of data. Relational databases use structured query language (SQL) to store and query data. This approach allows data to be stored in an easily understandable, consistent, and scalable way.

### Amazon Relational Database Service

A service that enables you to run relational databases in the AWS Cloud. Amazon RDS is a managed service that automates tasks such as hardware provisioning, database setup, patching, and backups. Many Amazon RDS database engines offer encryption at rest (protecting data while it is stored) and encryption in transit (protecting data while it is being sent and received). User can restrict access with security group & can plan for backup and recovery strategies. Multi AZ. You can reduce the load on your source DB Instance by routing read queries from your applications to the read replica.

AWS Multiple Availability Zone deployments allows for AWS to failover to a secondary database in case the primary one fails.

### Amazon Aurora

An enterprise-class relational database. It is faster than standard MySQL and PostgreSQL databases. Amazon Aurora helps to reduce your database costs by reducing unnecessary input/output (I/O) operations, while ensuring that your database resources remain reliable and available. Can auto-scale up to 128 TB.

Consider Amazon Aurora if your workloads require high availability. It replicates six copies of your data across three Availability Zones and continuously backs up your data to Amazon S3.

## Nonrelational databases

Place to create tables, where you can store and query data. Nonrelational databases are sometimes referred to as “NoSQL databases” because they use structures other than rows and columns to organize data. One type of structural approach is key-value pairs, where data is organized into items (keys), and items have attributes (values).

In a key-value database, you can add or remove attributes from items in the table at any time. Additionally, not every item in the table has to have the same attributes.

### Amazon DynamoDB

A key-value database service. It delivers single-digit millisecond performance at any scale. Is serverless, so you don’t have to provision, manage, install nor maintain any servers. DynamoDB automatically scales to adjust for changes in capacity while maintaining consistent performance. Dynamo DB chooses Performance over Consistency while delivering reads (“Eventually consistent”).

### Amazon Redshift

A data warehousing service that you can use for big data analytics. It offers the ability to collect data from many sources and helps you to understand relationships and trends across your data. Leader node: Receive queries and manage client connections.

## AWS Database Migration Service (AWS DMS)

Enables you to migrate relational databases, nonrelational databases, and other types of data stores.

With AWS DMS, you move data between a source database and a target database that can be of the same type or different types. During the migration, your source database remains operational, reducing downtime for any applications that rely on the database. Uses AWS Schema Conversion tool for heterogeneous database migrations where source database & destination database is different.

## Additional database services

* **DocumentDB:** Document database service that supports MongoDB workloads.
* **Neptune:** A graph database service. You can use Amazon Neptune to build and run applications that work with highly connected datasets
* **Quantum Ledger Database (Amazon QLDB):** Ledger database service. You can use it to review a complete history of all the changes that have been made to your application data.
* **Managed Blockchain:** A service that you can use to create and manage blockchain networks with open-source frameworks. **Blockchain** is a distributed ledger system that lets multiple parties run transactions and share data without a central authority.
* **ElastiCache:** A service that adds caching layers on top of your databases to help improve the read times of common requests.
* **DynamoDB Accelerator:** Is an in-memory cache for DynamoDB. It helps improve response times.

## Other services

**EMR:** is a cloud big data platform for processing vast amounts of data using open-source tools.

# Module 6: Security

## AWS shared responsibility model

AWS is responsible for the security of some parts of your environment and you (the customer) are responsible for other parts. It is divided into:

* **Customer responsibilities (“security in the cloud”):** Customers are responsible for the security of everything that they create and put in the AWS Cloud. The user is responsible for IAM roles and identities that can invoke the AWS Lambda functions. Client side data encryption, patch guest OS.

When using AWS services, you maintain complete control over the content. You are responsible for managing security requirements for your content, including which content you choose to store on AWS, which AWS services you use, and who has access to that content. You also control how access rights are granted, managed, and revoked.

* **AWS responsibilities (“security of the cloud”):** AWS operates, manages, and controls the components at all layers of infrastructure. This includes areas such as the host operating system, the virtualization layer, and even the physical security of the data centers from which services operate. AWS is responsible for protecting the global infrastructure that runs all of the services offered in the AWS Cloud. This infrastructure includes AWS Regions, Availability Zones, and edge locations.

### AWS Identity and Access Management (IAM)

Document that enables you to manage access to AWS services and resources securely. IAM gives you the flexibility to configure access based on your company’s specific operational and security needs. **It is free of cost**.

#### AWS account root user

When you first create an AWS account, you begin with an identity known as the root user. It has complete access to all the AWS services and resources in the account.

#### IAM users

An IAM user is an identity (represents the person or application) that you create in AWS. It consists of a name and credentials. By default, when you create a new IAM user in AWS, it has no permissions associated with it. To allow the IAM user to perform specific actions in AWS, you must grant the IAM user the necessary permissions.

#### IAM policies

Is a document that allows or denies permissions to AWS services and resources.

#### IAM groups

An IAM group is a collection of IAM users. When you assign an IAM policy to a group, all users in the group are granted permissions specified by the policy.

**IAM roles:** For getting temporary security credentials

**IAM tags:** Labels for custom attributes. Key-value pair.

### AWS Organizations

You can use AWS Organizations to consolidate and manage multiple AWS accounts within a central location. When you create an organization, AWS Organizations automatically creates a root, which is the parent container for all the accounts in your organization.

In AWS Organizations, you can centrally control permissions for the accounts in your organization by using **service control policies (SCPs)**. SCPs enable you to place restrictions on the AWS services, resources, and individual API actions that users and roles in each account can access.

SCPs can be used in Organizational Units and in an individual member account. Can be automated.

#### Organizational units

In AWS Organizations, you can group accounts into organizational units (OUs) to make it easier to manage accounts with similar business or security requirements. When you apply a policy to an OU, all the accounts in the OU automatically inherit the permissions specified in the policy. Can only have one parent.

## AWS Artifact

AWS Artifact is a service that provides on-demand access to AWS security and compliance reports and select online agreements. AWS Artifact consists of two main sections:

* **Agreements:** You can review, accept, and manage agreements for an individual account and for all your accounts in AWS Organizations. Different types of agreements are offered.
* **Reports:** Provide compliance reports from third-party auditors. These auditors have tested and verified that AWS is compliant with a variety of global, regional, and industry-specific security standards and regulations.

### Customer Compliance Center

Contains resources to help you learn more about AWS compliance.

## Denial-of-service attacks (DoS)

It is a deliberate attempt to make a website or application unavailable to users.

**Distributed denial-of-service attacks:**Multiple sources are used to start an attack that aims to make a website or application unavailable.

## AWS Shield

A service that protects applications against DDoS attacks. Route 53, Cloudfront and Load balancing also help in this. AWS Shield provides two levels of protection:

* **Standard:** Automatically protects all AWS customers at no cost. It protects your AWS resources from the most common, frequently occurring types of DDoS attacks. As network traffic comes into your applications, a variety of analysis techniques is used to detect malicious traffic in real time and automatically mitigates it.
* **Advanced:** Is a paid service that provides detailed attack diagnostics and the ability to detect and mitigate sophisticated DDoS attacks.

## AWS Key Management Service (AWS KMS)

Enables you to perform encryption operations through the use of cryptographic keys. You can use AWS KMS to create, manage, and use cryptographic keys. You can also control the use of keys across a wide range of services and in your applications. Uses envelope encryption while integrating with other AWS services.

A cryptographic key is a random string of digits used for locking (encrypting) and unlocking (decrypting) data.

With AWS KMS, you can choose the specific levels of access control that you need for your keys. For example, you can specify which IAM users and roles are able to manage keys. Alternatively, you can temporarily disable keys so that they are no longer in use by anyone. Your keys never leave AWS KMS, and you are always in control of them.

AWS managed KMS keys can only be rotated automatically as compared to the Customer Managed KMS keys that can be rotated automatically or manually.

## Amazon Inspector

Helps to improve the security and compliance of applications by running automated security assessments. It checks applications for security vulnerabilities and deviations from security best practices. After the Amazon Inspector has performed an assessment, it provides you with a list of security findings.

Amazon Inspector provides two types of packages. Network reachability rules package checks network accessibility checks on Amazon EC2 instance. Host assessment rules package checks vulnerabilities on Amazon EC2 instances. Uses predefined templates.

## Amazon GuardDuty

It is a service that provides intelligent threat detection for your AWS infrastructure and resources. It identifies threats by continuously monitoring the network activity and account behavior within your AWS environment, including S3.

## Other services

**AWS WAF:** It is a web application firewall that lets you monitor network requests that come into your web applications.

**Bastion Host Overview:** Bastion means a structure for Fortification to protect things behind it. In AWS, a Bastion host can be used to securely access instances in the private subnets. They need to be created in multiple AZ for redundancy.

**Certificate Manager:** Allows the web administrator to maintain one or several SSL/TLS certificates. To verify one: a CNAME record would need to be created and the administrator would need to acknowledge a verification email.

**Cloud HSM:** Allows for the administrator to have full and exclusive control over the generation and management cryptographic keys on actual hardware security modules that are physically stored in AWS data centers.

**Cognito:** Lets you add user sign-up, sign-in, and access control to your web and mobile apps quickly and easily. Supports sign-in with social identity providers, such as Apple, Facebook, Google, and other login services.

**Encryption SDK:** Improves implementation of the client-side encryption.

**Firewall manager:** Can manage VPC security groups, AWS Shield Advanced and WAF rules on one platform even across multiple AWS accounts. Since the DB does not require a Public IP, it is always a good practice to host the DB server on a non-default subnet that does not allocate a public IP by default. Also, the DB server should not have a route to the internet gateway.

**Identity Federation:** So users can access the AWS environment using their active directory credentials. It is possible for users to log into the AWS environment using their Facebook, Twitter or LinkedIn login credentials.

**Secrets Manager:** Easy way to safely store encrypted credentials and perform on demand retrieval.

**Security hub:** Facilitates the view of high priority security alerts. provides a view of across-account security status and gives security alerts.

**Security Token service:** web service that enables you to request temporary, limited-privilege credentials for AWS IAM users or for users that you authenticate (federated users). Mobile apps need access to AWS resources. STS + Identity federation = Session

**Single Sign-On (SSO):** Simplifies access management and manages access to multiple AWS accounts and business applications centrally.

# Module 7: Monitoring and Analytics

## Amazon CloudWatch

Amazon CloudWatch is a web service that enables you to monitor and manage various metrics (CPU, RAM) and configure alarm actions based on data from those metrics. CloudWatch uses metrics to represent the data points for your resources in graphs.

### CloudWatch alarms

With CloudWatch, you can create alarms that automatically perform actions if the value of your metric has gone above or below a predefined threshold.

### CloudWatch dashboard

The CloudWatch dashboard feature enables you to access all the metrics for your resources from a single location.

## AWS CloudTrail

Records API calls for your account (and from other accounts). The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, and more. Events are typically updated in CloudTrail within 15 minutes after an API call.

**CloudTrail Insights:** This optional feature allows CloudTrail to automatically detect unusual API activities in your AWS account.

## AWS Trusted Advisor

A web service that inspects your AWS environment and provides real-time recommendations in accordance with AWS best practices. The inspection includes security checks, such as Amazon S3 buckets with open access permissions.

Trusted Advisor compares its findings to AWS best practices in five categories: cost optimization, performance, security, fault tolerance, and service limits. For the checks in each category, Trusted Advisor offers a list of recommended actions and additional resources to learn more about AWS best practices.

**AWS Trusted Advisor Notification** is an optional service that needs to be set up from the dashboard.

### AWS Trusted Advisor dashboard

Where you can review completed checks for cost optimization, performance, security, fault tolerance, and service limits. For each category:

* Green 🡪 no problems
* Orange 🡪 recommended investigations
* Red 🡪 recommended actions.

**Penetration testing** can be performed by customers, provided they work with the list of services mentioned by AWS.

## Other services

**Change Sets:** Preview changes to resources when a stack is executed.

**Compute optimizer:** ML based tool that analyses metrics of historic utilization and makes recommendations of compute services.

**Config:** is a service that enables you to assess, audit, and evaluate the configurations of your AWS resources.

**Detective:** makes it easy to analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities. Uses machine learning.

**Disaster recovery techniques:** From highest to lowest highest downtime: Backup & Restore, Pilot light, Warm standby, Multi site.

**OpsWorks:** is a configuration management service that provides managed instances of Chef and Puppet. Chef and Puppet are automation platforms that allow you to use code to automate the configurations of your servers. These platforms will allow for the use of code to automate the configurations of EC2 instances. Mainly used for Linux and public endpoints.

**Quick Sight:** Graphic reports. Scalable, serverless, embeddable, machine learning-powered business intelligence (BI) service that lets you create BI dashboards.

**Service health dashboard:** most up-to-the-minute information on AWS service availability here.

**Well-Architected tool:** Provides advice on architecting the workload in the cloud.

# Module 8: Pricing and support

## AWS Free Tier

The AWS Free Tier enables you to begin using certain services without having to worry about incurring costs for the specified period.

* **Always Free:** These offers do not expire and are available to all AWS customers. For example, AWS Lambda allows 1 million free requests and up to 3.2 million seconds of compute time per month. Amazon DynamoDB allows 25 GB of free storage per month.
* **12 Months Free:** These offers are free for 12 months following your initial sign-up date to AWS.
* **Trials:** Short-term free trial offers start from the date you activate a particular service. The length of each trial might vary by number of days or the amount of usage in the service.

## How AWS pricing works

* **Pay for what you use:** For each service, you pay for exactly the number of resources that you use.
* **Pay less when you reserve:** Some services offer reservation options that provide a significant discount compared to On-Demand Instance pricing.
* **Pay less with volume-based discounts when you use more:** Some services offer tiered pricing, so the per-unit cost is incrementally lower with increased usage. For example, the more Amazon S3 storage space you use, the less you pay for it per GB.
* **No cost for inbound data transfer**
* **Saving plans:** 72% (not available in all regions)

## AWS Pricing Calculator

It lets you explore AWS services and create an estimate for the cost of your use cases on AWS. You can organize your AWS estimates by groups that you define. A group can reflect how your company is organized, such as providing estimates by cost center.

### AWS pricing examples

***AWS Lambda:*** you are charged based on the number of requests for your functions and the time that it takes to run.

***Amazon S3:***

* **Storage:** You pay for only the storage that you use.
* **Requests and data retrievals:** You pay for requests made to your Amazon S3 objects and buckets.
* **Data transfer:** There is no cost to transfer data between different Amazon S3 buckets or from Amazon S3 to other services within the same AWS Region.
* **Management and replication:** You pay for the storage management features that you have enabled on your account’s Amazon S3 buckets. These features include Amazon S3 inventory, analytics, and object tagging.

## AWS Billing Dashboard

Where you can pay your AWS bill, monitor your usage, and analyze and control your costs.

* Compare your current month-to-date balance with the previous month and get a forecast of the next month based on current usage.
* View month-to-date spend by service.
* View Free Tier usage by service.
* Access Cost Explorer and create budgets.
* Purchase and manage Savings Plans.
* Publish AWS Cost and Usage Reports.

## Consolidated billing

The consolidated billing feature of AWS Organizations enables you to receive a single bill for all AWS accounts in your organization. By consolidating, you can easily track the combined costs of all the linked accounts in your organization. The default maximum number of accounts allowed for an organization is 4, but you can contact AWS Support to increase your quota, if needed.

Also, you can share bulk for discount pricing, like Savings Plans, and Reserved Instances.

## AWS Budgets

You can create budgets to plan your service usage, service costs, and instance reservations.

The information in AWS Budgets updates three times a day. This helps you to accurately determine how close your usage is to your budget. You can also set custom alerts when your usage exceeds (or is forecasted to exceed) the budgeted amount.

## AWS Cost Explorer

It is a tool that enables you to visualize, understand, and manage your AWS costs and usage over time. It includes a default report of the costs and usage for your top five cost-accruing AWS services. Includes data for up to the last 12 months

Identifies areas that need further inquiry and provides trends that you can use to understand costs. Can provide usage-based forecasts of estimated billing costs and usage for the coming 12 months.

## AWS Support

### Basic Support

It is free for all AWS customers. It includes access to whitepapers, documentation, and support communities. You can also contact AWS for billing questions and service limit increases.

With Basic Support, you have access to a limited selection of AWS Trusted Advisor checks. Additionally, you can use the **AWS Personal Health Dashboard**, a tool that provides alerts and remediation guidance when AWS is experiencing events that may affect you (can be used to identify an issue with the Amazon CloudWatch service) (Availability and performance of AWS services, alerts and notifications).

### Developer, Business, and Enterprise Support

They include all the benefits of Basic Support, in addition to the ability to open an unrestricted number of technical support cases. These three Support plans have pay-by-the-month pricing and require no long-term contracts.

* Developer:
  + Best practice guidance
  + Client-side diagnostic tools
  + Building-block architecture support
* Business Support:
  + Use-case guidance to identify AWS offerings, features, and services
  + All AWS Trusted Advisor checks
  + Limited support for third-party software
* Enterprise
  + Application architecture guidance
  + Infrastructure event management: Provides use-case, architectural and scaling guidance.

You can see the full table of differences in the [official documentation](https://aws.amazon.com/premiumsupport/plans/).

## Technical Account Manager (TAM)

Included with Enterprise Support. Primary point of contact at AWS. They provide guidance, architectural reviews, and ongoing communication with your company as you plan, deploy, and optimize your applications. They help you design solutions that efficiently use multiple services together through an integrated approach.

## AWS Marketplace

AWS Marketplace is a digital catalog that includes thousands of software listings from independent software vendors. You can use AWS Marketplace to find, test, and buy software that runs on AWS.

CPU bound SW licenses require a dedicated Host 🡪 expensive. High availability may involve manual recovery of resources.

## Other services

**Total cost of ownership:** Number of servers migrated to AWS. The TCO calculator will make suggestions and recommendations on appropriate resource types based on the user’s input values and settings.

**TCO Calculator:** Use this calculator to compare the cost of running your applications in an on-premises or colocation environment to AWS.

# 

# Module 9: Migration and innovation

## Six core perspectives of the Cloud Adoption Framework

Organizes guidance into six areas of focus, called Perspectives. Each Perspective addresses distinct responsibilities.

1. **Business:** Ensures that IT aligns with business needs.
2. **People:** Evaluate organizational structures and roles and process requirements. This helps prioritize training and staffing.
3. **Governance:** To understand how to update the staff skills and processes to ensure business governance in the cloud.
4. **Platform:** Includes principles and patterns for implementing new solutions on the cloud and migrating on-premises workloads to the cloud.
5. **Security:** Ensures that the organization meets security objectives for visibility, auditability, control, and agility.
6. **Operations:** Align with and support the operations of the business.

## Six strategies for migration

1. **Rehosting:** Involves moving applications without changes to another host.
2. **Replatforming:** involves making a few cloud optimizations to realize a tangible benefit.
3. **Refactoring:** involves reimagining how an application is architected and developed by using cloud-native features.
4. **Repurchasing:** involves moving from a traditional license to a software-as-a-service model.
5. **Retaining:** consists of keeping applications that are critical for the business in the source environment.
6. **Retiring:** process of removing applications that are no longer needed.

## AWS Snow Family members

It is a collection of physical devices that help to physically transport data into and out of AWS.

* **AWS Snowcone** is a small, rugged, and secure edge computing and data transfer device.
  + It features 2 CPUs, 4 GB of memory, and 8 TB of usable storage.
* **AWS Snowball**
  + **Snowball Edge Storage Optimized** devices are well suited for large-scale data migrations.
    - Storage: 80 TB - 1 TB
    - Compute: 40 vCPUs, and 80 GiB of memory.
  + **Snowball Edge Compute Optimized** provides powerful computing resources.
    - Storage: 42-TB - 7.68 TB
    - Compute: 52 vCPUs, 208 GiB of memory, and an optional NVIDIA Tesla V100 GPU.
* **AWS Snowmobile** is a data transfer service used to move large amounts of data to AWS. Up to 100 petabytes.

## Innovation

**Serverless applications:** Refers to applications that don’t require you to provision, maintain, or administer servers. You don’t need to worry about fault tolerance or availability.

**Artificial intelligence:**

* **Amazon Transcribe:** Convert speech to text
* **Amazon Comprehend:** Discover patterns in text
* **Amazon Fraud Detector:** Identify potentially fraudulent online activities
* **Amazon Lex:** Build voice and text chatbots
* **Amazon Augmented AI:** provides built-in human review workflows for common machine learning use cases.

**Machine learning:** AWS offers Amazon **SageMaker** to remove the difficult work from the process and empower you to build, train, and deploy ML models quickly.

**Macie:** Fully managed service that provides data security and privacy using machine learning algorithms. For S3. Can be used to detect users' personal credit card numbers from data stored in Amazon S3. Discover, Classify, Protect.

**Amazon Textract:** A fully managed machine learning service that automatically extracts printed text, handwriting, and other data from scanned documents

## Other services

**DataSync:** To move huge amounts of data (hundreds of terabytes) between on-prem storage to S3, EFS, FSx. Can work over AWS Direct Connect or not. It is not an agentless data transfer service.

# Module 10: Cloud journey

## The AWS Well-Architected Framework

It helps you understand how to design and operate reliable, secure, efficient, and cost-effective systems in the AWS Cloud. It provides a way for you to consistently measure your architecture against best practices and design principles and identify areas for improvement. It is based on five pillars:

1. **Operational excellence:** The ability to run and monitor systems to deliver business value and to continually improve supporting processes and procedures. They include performing operations as code, annotating documentation, anticipating failure, and frequently making small, reversible changes.
2. **Security:** The ability to protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies. They include implementing a strong identity foundation, enable traceability, apply security at all layers, automate security best practices, protect data in transit and at rest, keep people away from data, prepare for security events.
3. **Reliability:** Mainly to recover from failures. The ability of a system to:
   1. Recover from infrastructure or service disruptions
   2. Dynamically acquire computing resources to meet demand
   3. Mitigate disruptions such as misconfigurations or transient network issues
4. **Performance efficiency:** The ability to use computing resources efficiently to meet system requirements
5. **Cost optimization:** the ability to run systems to deliver business value at the lowest price point. To manage users:
   1. Use of AWS Organizations with respective OUs that differentiate billing across the company’s functions.
   2. Implement the most stringent security measures on the VPC-edge rather than on the resource hosts.

You can find more information about the pillars in the official documentation:

* AWS Well-Architected [[Link](https://aws.amazon.com/architecture/well-architected/?wa-lens-whitepapers.sort-by=item.additionalFields.sortDate&wa-lens-whitepapers.sort-order=desc)]
* The 5 Pillars of the AWS Well-Architected Framework [[Link](https://aws.amazon.com/blogs/apn/the-5-pillars-of-the-aws-well-architected-framework/)]

## Advantages of cloud computing

1. Trade upfront expense for variable expense: Instead of investing heavily, you can pay only when you consume.
2. Benefit from massive economies of scale: Because usage from hundreds of thousands of customers, providers can achieve higher economies of scale.
3. Stop guessing capacity: You don’t have to predict how much infrastructure capacity you will need
4. Increase speed and agility: The flexibility makes it easier to develop and deploy applications.
5. Stop spending money running and maintaining data centers
6. Go global in minutes