[AWS Ramp-Up Guide: Cloud Essentials](#)
https://www.examtopics.com/exams/amazon/aws-certified-cloud-practitioner/view/

AWS Essentials

Module 1: Intro AWS

1. client-server model

In computing, 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.

2. Deployment models for cloud computing

The three cloud computing deployment models are **cloud-based, on-premises, and hybrid.**

On-premises deployment is also known as a private cloud deployment. In this model, resources are deployed on-premises using virtualization and resource management tools.

In a hybrid deployment, cloud-based resources are connected to on-premises infrastructure. You might want to use this approach in several situations. For example, you might have legacy applications that are better-maintained on-premises, or government regulations require your business to keep certain records on-premises.

Module 2: compute in the cloud

Elastic Compute Cloud (EC2)

a virtual server to run applications in the AWS Cloud.

1. Launch

First, you launch an instance. You begin by selecting a template with basic configurations for your instance, including the operating system, application server, and applications. You also select the instance type, which is the specific hardware configuration of your instance.

2. Connect

Next, connect to the instance. You can connect to the instance in several ways. Your programs and applications have multiple methods to connect directly to the instance and exchange data. Users can also connect to the instance by logging in and accessing the computer desktop.

3. Use

You can run commands to install software, add storage, copy and organize files, and more.

EC2 instance types

1. **General purpose** instances provide a balance of compute, memory, and networking resources.

2. **Compute optimized** instances are ideal for compute-bound applications that benefit from *high-performance processors*. ideal for high-performance web servers, compute-intensive applications servers, and dedicated gaming servers. You can also use compute optimized instances for *batch processing* workloads that require processing many transactions in a single group.

3. **Memory optimized** instances are designed to deliver fast performance for workloads that process large datasets in memory. In computing, memory is a temporary storage area. It holds all the data and instructions that a central processing unit (CPU) needs to be able to complete actions. Before a computer program or application is able to run, it is loaded from storage into memory. This preloading process gives the CPU direct access to the computer program. Suppose that you have a workload that requires *large amounts of data* to be preloaded before running an application. This scenario might be a *high-performance database* or a workload that involves performing *real-time processing* of a large amount of unstructured data.

4. **Accelerated computing** instances use hardware accelerators, or coprocessors, to perform some functions more efficiently than is possible in software running on CPUs. Examples of these functions include floating-point number calculations, graphics processing, and data pattern matching. graphics applications, game streaming, and application streaming.

5. **Storage optimized** instances are designed for workloads that require high, sequential read and write access to large datasets on local storage. Examples of workloads suitable for storage optimized instances include distributed file systems, data warehousing applications, and high-frequency online transaction processing (OLTP) systems. In computing, the term input/output operations per second (IOPS) is a metric that measures the performance of a storage device. It indicates how many different input or output operations a device can perform in one second. Storage optimized instances are designed to deliver tens of thousands of low-latency, random IOPS to applications.

EC2 pricing

On-Demand Instances are 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.

Reserved Instances are a billing discount applied to the use of On-Demand Instances in your account.1-year or 3-year

Standard Reserved Instances: This option is a good fit if you know the EC2 instance type and size you need for your steady-state applications and in which AWS Region you plan to run them.
Convertible Reserved Instances: If you need to run your EC2 instances in different Availability Zones or different instance types, then Convertible Reserved Instances might be right for you. Note: You trade in a deeper discount when you require flexibility to run your EC2 instances.

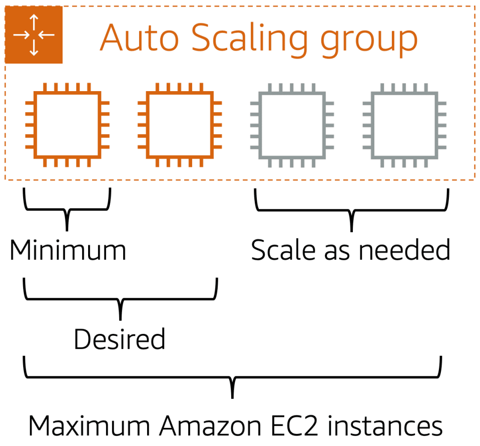
EC2 Instance Savings Plans reduce your EC2 instance costs when you make an hourly spend commitment to an instance family and Region for a 1-year or 3-year term.

Spot Instances are ideal for workloads with flexible start and end times, or that can withstand interruptions. Spot Instances use unused Amazon EC2 computing capacity and offer you cost savings at up to 90% off of On-Demand prices.Suppose that you have a background processing job that can start and stop as needed (such as the data processing job for a customer survey)

Dedicated Hosts are physical servers with Amazon EC2 instance capacity that is fully dedicated to your use.
You can use your existing per-socket, per-core, or per-VM software licenses to help maintain license compliance.

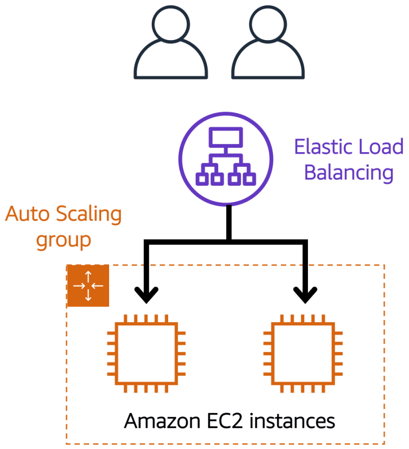
EC2 Scaling

Amazon EC2 Auto Scaling enables you to automatically add or remove Amazon EC2 instances in response to changing application demand.
Dynamic scaling responds to changing demand.
Predictive scaling automatically schedules the right number of Amazon EC2 instances based on predicted demand.
The third configuration that you can set in an Auto Scaling group is the maximum capacity. For example, you might configure the Auto Scaling group to scale out in response to increased demand, but only to a maximum of four Amazon EC2 instances.



Directing Traffic with Elastic Load Balancing

Elastic Load Balancing is the AWS service that automatically distributes incoming application traffic across multiple resources, such as Amazon EC2 instances



Messaging and Queuing

Monolithic applications vs microservices
Applications are made of multiple components. The components communicate with each other to transmit data, fulfill requests, and keep the application running. These components might include databases, servers, the user interface, business logic, and so on.

In a microservices approach, application components are loosely coupled.

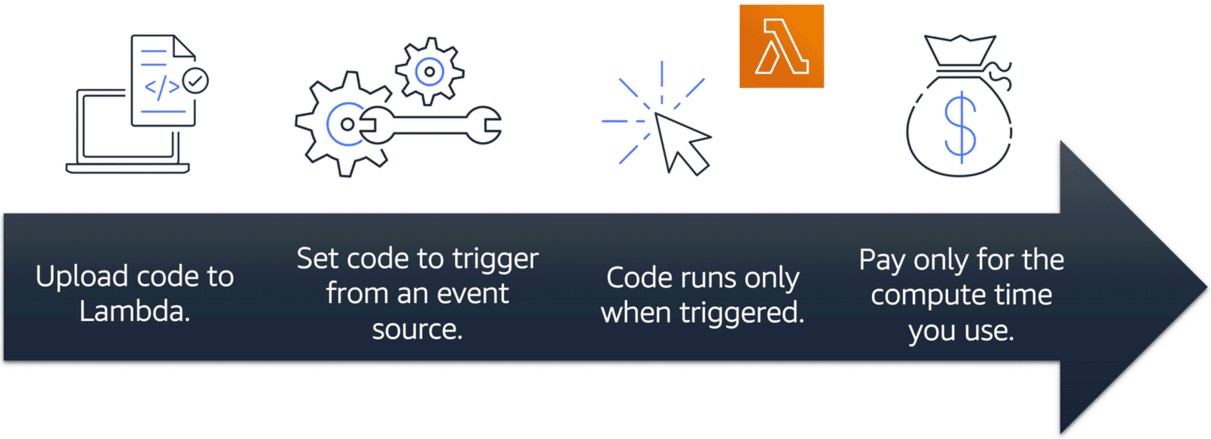
Two services facilitate application integration: Amazon Simple Notification Service (Amazon SNS) and Amazon Simple Queue Service (Amazon SQS).

Amazon SNS is a publish/subscribe service. Using Amazon SNS topics, a publisher publishes messages to subscribers.
In Amazon SNS, subscribers can be web servers, email addresses, AWS Lambda functions, or several other options.
Example is newsletter: Publishing updates from a single topic vs Publishing updates from multiple topics

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.
Example: Fulfilling an order (customer cashier barista with paper) vs Orders in a queue (customer cashier QUEUE barista)

Serverless computing

The term “serverless” means that your code runs on servers, but you do not need to provision or manage these servers.
Another benefit of serverless computing 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.
An AWS service for serverless computing is **AWS Lambda**. Lambda is a service that lets you run code without needing to provision or manage servers. For example, a simple Lambda function might involve automatically resizing uploaded images to the AWS Cloud. In this case, the function triggers when uploading a new image.


Upload code to Lambda.
Set code to trigger from an event source.
Code runs only when triggered.
Pay only for the compute time you use.

Containers

Containers provide you with a standard way to package your application's code and dependencies into a single object. You can also use containers for processes and workflows in which there are essential requirements for security, reliability, and scalability.

Amazon Elastic Container Service **(Amazon ECS)** is a highly scalable, high-performance container management system that enables you to run and scale containerized applications on AWS. Amazon ECS supports Docker containers.

Amazon Elastic Kubernetes Service **(Amazon EKS)** is a fully managed service that you can use to run Kubernetes on AWS.

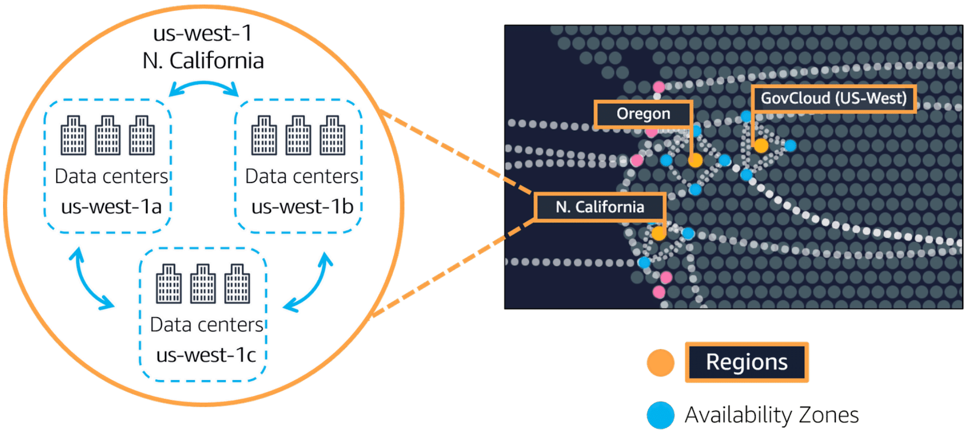
AWS Fargate is a serverless compute engine for containers. It works with both Amazon ECS and Amazon EKS. When using AWS Fargate, you do not need to provision or manage servers. AWS Fargate manages your server infrastructure for you.

Module 3: Global infra & Reliability

Selecting a Region

1. Compliance with data governance and legal requirements
2. Proximity to your customers
3. Available services within a Region
4. Pricing

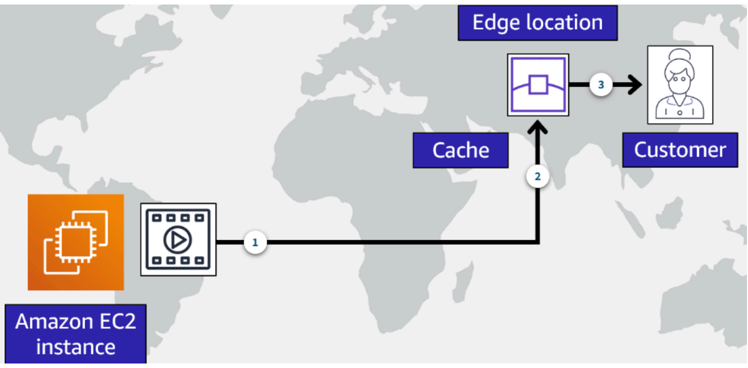
Availability Zones



An Availability Zone is a single data center or a group of data centers within a Region. This is close enough to have low latency (the time between when content requested and received) between Availability Zones. However, if a disaster occurs in one part of the Region, they are distant enough to reduce the chance that multiple Availability Zones are affected.

Edge locations

An edge location is a site that Amazon **CloudFront** uses to store cached copies of your content closer to your customers for faster delivery.



Amazon **CloudFront** is a content delivery service. It uses a network of edge locations to cache content and deliver content to customers all over the world. When content is cached, it is stored locally as a copy. This content might be video files, photos, webpages, and so on.

AWS **Outposts** is a fully managed service that extends AWS infrastructure, services, APIs, and tools to customer premises.

Ways to interact with AWS services

1. **AWS Management Console** is a web-based interface for accessing and managing AWS services. You can quickly access recently used services and search for other services by name, keyword, or acronym. The console includes wizards and automated workflows that can simplify the process of completing tasks. 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.

2. To save time when making API requests, you can use the **AWS Command Line Interface (AWS CLI).** AWS CLI enables you to control multiple AWS services directly from the command line within one tool. By using AWS CLI, you can automate the actions that your services and applications perform through scripts. For example, you can use commands to launch an Amazon EC2 instance, connect an Amazon EC2 instance to a specific Auto Scaling group, and more.

3. **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.

4. With **AWS Elastic Beanstalk** (Platform as a Service (PaaS), you provide code and configuration settings, and Elastic Beanstalk deploys the *resources necessary* to perform the following tasks:
    a. Adjust capacity
    b. Load balancing
    c. Automatic scaling
    d. Application health monitoring

5. With **AWS CloudFormation** Infrastructure as Code (IaC), you can treat your infrastructure as code. This means that you can build an environment by writing lines of code instead of using the AWS Management Console to individually provision resources. 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. It determines the right operations to perform when managing your stack and rolls back changes automatically if it detects errors. Uses templates (JSON or YAML) to describe infrastructure

Choosing Between the Two
Use Elastic Beanstalk if you want a simple, automated way to deploy applications without managing infrastructure details.
Choose CloudFormation if you need precise control over your AWS resources and want to manage infrastructure as code.

To recap, the AWS Management Console is great for learning and providing a visual for the user. The AWS Management Console is a manual tool. So right off the bat, it isn't a great option for automation. You can instead use the CLI to script your interactions with AWS using the terminal. You can use the SDKs to write programs to interact with AWS for you or you can use manage tools like AWS Elastic Beanstalk or AWS CloudFormation.

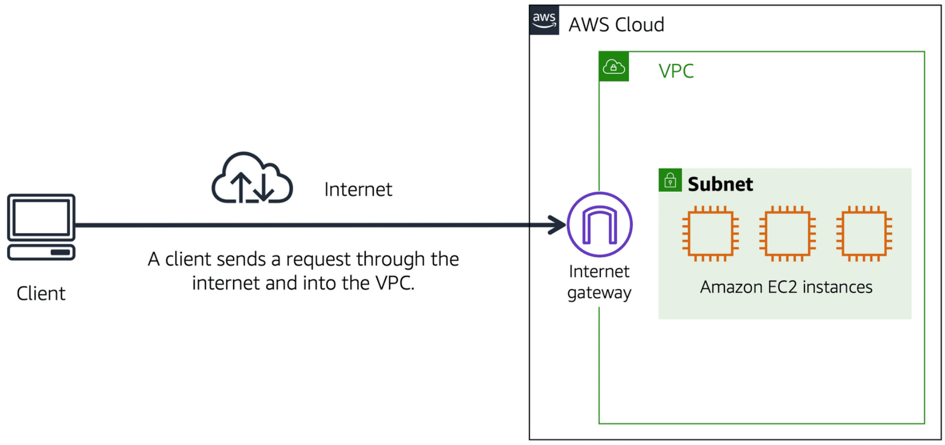
Module 4: Networking

Virtual Private Cloud (Amazon VPC)

A networking service that you can use to establish boundaries around your AWS resources is a VPC. 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 virtual private cloud (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.

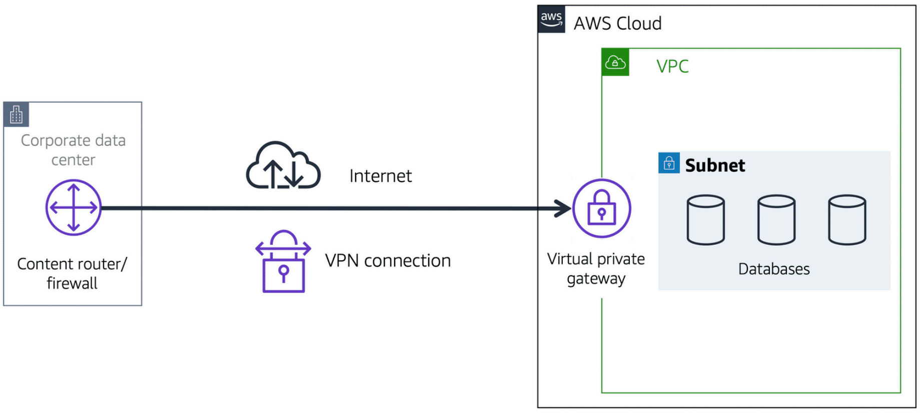
Internet gateway

To **allow public traffic** from the internet to access your VPC, you attach an internet gateway to the VPC. An internet gateway is a connection between a VPC and the internet. You can think of an internet gateway as being similar to a doorway that customers use to enter the coffee shop. Without an internet gateway, no one can access the resources within your VPC.



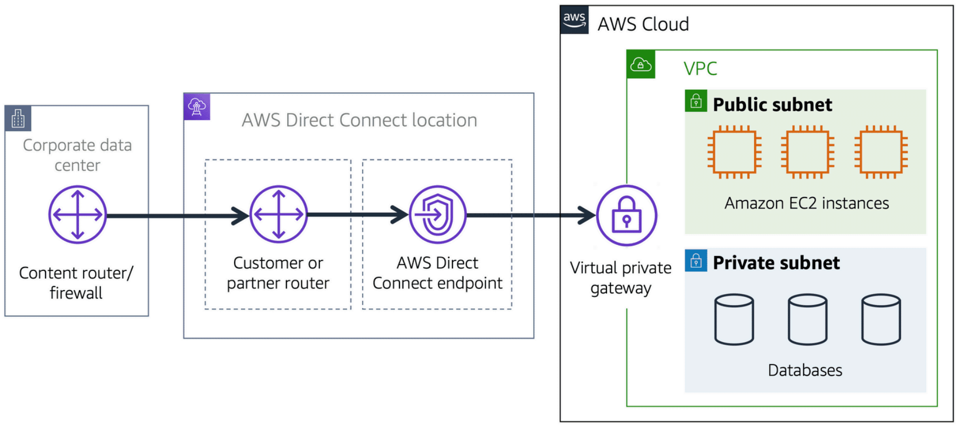
Virtual private gateway

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



AWS Direct Connect

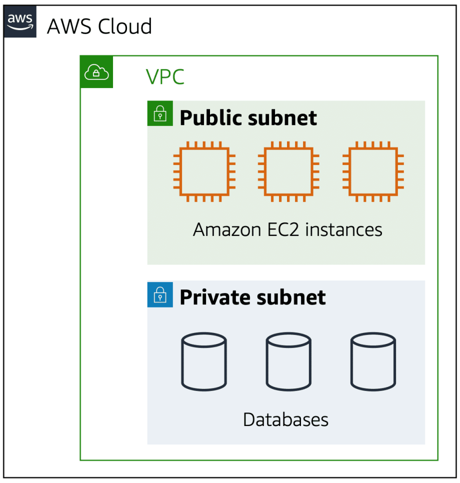
AWS Direct Connect is a service that lets you to establish a dedicated private connection between your data center and a VPC.



Subnets and Network Access Control Lists

Subnets

A subnet is 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. For example, you might have an application that involves Amazon EC2 instances in a public subnet communicating with databases that are located in a private subnet.

Network traffic in a VPC

When a customer requests data from an application hosted in the AWS Cloud, this request is sent as a **packet**. A packet is a unit of data sent over the internet or a network.

It enters into a VPC through an **internet gateway**. Before a packet can enter into a subnet or exit from a subnet, 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)**.

Network ACLs

A network ACL is a 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 AC*L allows all inbound and outbound traffic*, but you can modify it by adding your own rules. 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.
After a packet has entered a subnet, it must have its permissions evaluated for resources within the subnet, such as Amazon EC2 instances.

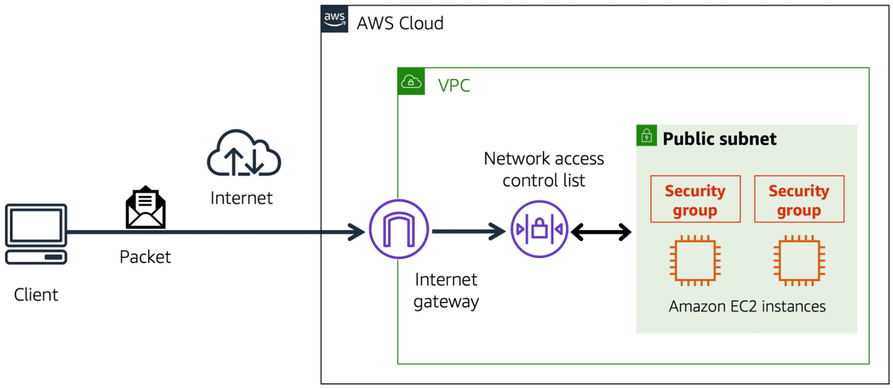
The VPC component that checks packet permissions for an Amazon EC2 instance is a security group.

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 should be allowed; any other traffic would then be denied. If you have multiple Amazon EC2 instances within the same VPC, you can associate them with the same security group or use different security groups for each instance.

Stateful packet filtering

Security groups perform stateful packet filtering. They remember previous decisions made for incoming packets. When a packet response for that request returns to the instance, the security group remembers your previous request. The security group allows the response to proceed, regardless of inbound security group rules.



Security Group is Stateful (remembers for outband)- Network ACL is Stateless
Security groups deny all inbound traffic by default.
network ACL allows all inbound and outbound traffic by default

VPC component recall

1. Private subnet: A network segment within a virtual private cloud (VPC) that is not directly accessible from the public internet. Like databases containing customer info
2. Virtual private gateway: create a VPN connection between VPC & internal corporate network.
3. Public subnet: support customer facing website.
4. AWS Direct Connect: establish a dedicated, private connectivity between on-premises data center network & the VPC.

Global Networking

Domain Name System (DNS)

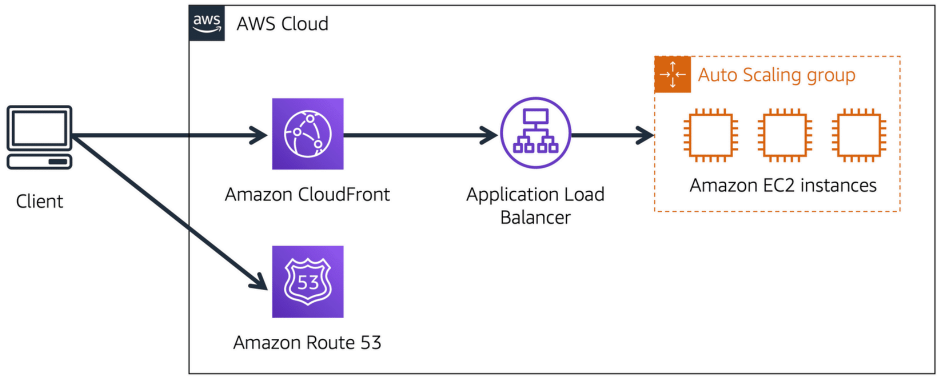
Suppose that AnyCompany has a website hosted in the AWS Cloud. Customers enter the web address into their browser, and they are able to access the website. This happens because of Domain Name System (DNS) resolution. DNS resolution involves a customer DNS resolver communicating with a company DNS server.

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 (such as Amazon EC2 instances and load balancers). 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 of your domain names within a single location.



Suppose that AnyCompany's application is running on several Amazon EC2 instances. These instances are in an Auto Scaling group that attaches to an Application Load Balancer.

1. A customer requests data from the application by going to AnyCompany's website.
2. Amazon Route 53 uses DNS resolution to identify AnyCompany.com's corresponding IP address, 192.0.2.0. This information is sent back to the customer.
3. The customer's request is sent to the nearest edge location through Amazon CloudFront.
4. Amazon CloudFront connects to the Application Load Balancer, which sends the incoming packet to an Amazon EC2 instance.

- A **public subnet** is a section of a VPC that contains public-facing resources.
- An **edge location** is a site that Amazon CloudFront uses to store cached copies of your content for faster delivery to customers.
- A **security group** is a virtual firewall that controls inbound and outbound traffic for an Amazon EC2 instance.
- **Internet gateway** is used to connect a VPC to the internet
- AWS Direct Connect is a service that enables you to establish a dedicated private connection between your data center and VPC.

Module 5: Storage & Databases

Instance Stores and Amazon Elastic Block Store (Amazon EBS)

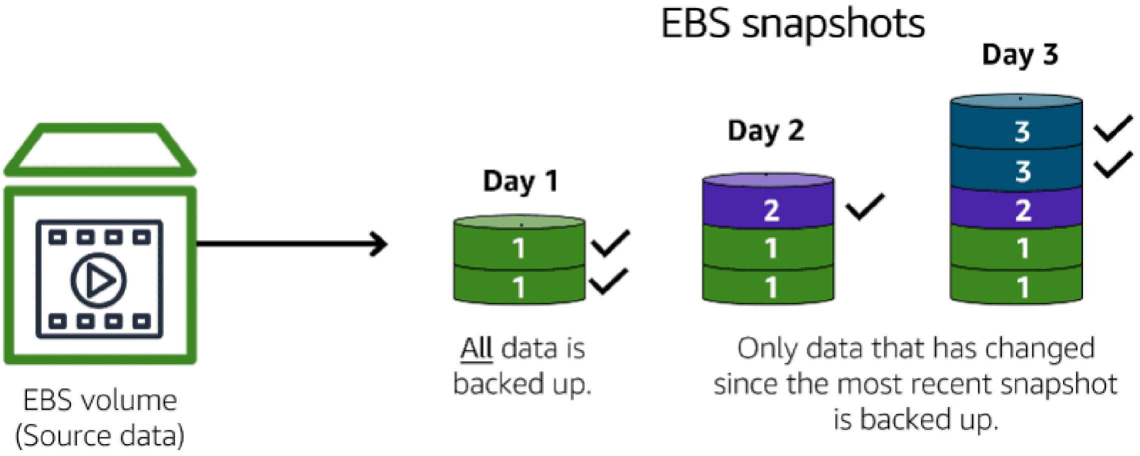
Instance stores
Block-level storage volumes behave like physical hard drives.

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 that you can use with Amazon EC2 instances. 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, it can attach to an Amazon EC2 instance.

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.



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.

Incremental backups are different from full backups, in which all the data in a storage volume copies each time a backup occurs. The full backup includes data that has not changed since the most recent backup.

EBS characteristics:

- Best for data that requires retention
- Separate drives from the host computer of an EC2 instance

Simple Storage Service (Amazon S3)

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. An object's key is its unique identifier.

- when you modify a file in block storage, only the pieces that are changed are updated. When a file in object storage is modified, the entire object is updated.

Amazon Simple Storage Service (Amazon S3) is a service that provides object-level storage. Amazon S3 stores data as objects in buckets.

You can upload any type of file to Amazon S3, such as images, videos, text files, and so on. For example, you might use Amazon S3 to store backup files, media files for a website, or archived documents. Amazon S3 offers unlimited storage space. The maximum file size for an object in Amazon S3 is 5 TB.

Amazon S3 storage classes

With Amazon S3, you pay only for what you use.

- How often you plan to retrieve your data
- How available you need your data to be

Key factors to consider: access frequency, availability needs, retrieval speed, and cost.

1. S3 Standard
    - Designed for frequently accessed data
    - Stores data in a minimum of three Availability Zones

Amazon S3 Standard provides high availability for objects. This makes it a good choice for a wide range of use cases, such as websites, content distribution, and data analytics. Amazon S3 Standard has a higher cost than other storage classes intended for infrequently accessed data and archival storage.

2. S3 Standard-Infrequent Access (S3 Standard-IA)

    - Ideal for infrequently accessed data
    - Similar to Amazon S3 Standard but has a lower storage price and higher retrieval price

Amazon S3 Standard-IA is ideal for data infrequently accessed but requires high availability when needed. Both Amazon S3 Standard and Amazon S3 Standard-IA store data in a minimum of three Availability Zones. Amazon S3 Standard-IA provides the same level of availability as Amazon S3 Standard but with a lower storage price and a higher retrieval price.

3. S3 One Zone-Infrequent Access (S3 One Zone-IA)
    - Stores data in a single Availability Zone
    - Has a lower storage price than Amazon S3 Standard-IA

Compared to S3 Standard and S3 Standard-IA, which store data in a minimum of three Availability Zones, S3 One Zone-IA stores data in a single Availability Zone. This makes it a good storage class to consider if the following conditions apply:

- You want to save costs on storage.
- You can easily reproduce your data in the event of an Availability Zone failure.

4. S3 Intelligent-Tiering
    - Ideal for data with unknown or changing access patterns
    - Requires a small monthly **monitoring** and automation fee per object

In the S3 Intelligent-Tiering storage class, **Amazon S3 monitors objects' access patterns.** 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.

5. S3 Glacier Instant Retrieval
    - Works well for archived data that requires immediate access
    - Can retrieve objects within a few milliseconds

When you decide between the options for archival storage, consider how quickly you must retrieve the archived objects. You can retrieve objects stored in the S3 Glacier Instant Retrieval storage class within milliseconds, with the same performance as S3 Standard.

6. S3 Glacier Flexible Retrieval
    - Low-cost storage designed for data archiving
    - Able to retrieve objects within a few minutes to hours

S3 Glacier Flexible Retrieval is a low-cost storage class that is ideal for data archiving. For example, you might use this storage class to store archived customer records or older photos and video files. You can retrieve your data from S3 Glacier Flexible Retrieval from **1 minute to 12 hours**.

7. S3 Glacier Deep Archive

- Lowest-cost object storage class ideal for archiving
- Able to retrieve objects within **12 hours**

S3 Deep Archive supports long-term retention and digital preservation for data that might be accessed once or twice in a year. This storage class is the lowest-cost storage in the AWS Cloud, with data retrieval from 12 to 48 hours. All objects from this storage class are replicated and stored across at least three geographically dispersed Availability Zones.

8. S3 Outposts

- Creates S3 buckets on Amazon S3 Outposts
- Makes it easier to retrieve, store, and access data on AWS Outposts

Amazon S3 Outposts delivers object storage to your on-premises AWS Outposts environment. Amazon S3 Outposts is designed to store data durably and redundantly across multiple devices and servers on your Outposts. It works well for workloads with local data residency requirements that must satisfy demanding performance needs by keeping data close to on-premises applications.

1. S3 Standard: Frequent access, high availability, 3+ AZs, higher cost
2. S3 Standard-IA: Infrequent access, high availability, 3+ AZs, lower storage cost, higher retrieval cost
3. S3 One Zone-IA: Infrequent access, single AZ, lowest cost for IA, less durability
4. S3 Intelligent-Tiering: Unknown/changing access patterns, auto-tiering, small monitoring fee
5. S3 Glacier Instant Retrieval: Archived data, immediate access (milliseconds), higher retrieval cost
6. S3 Glacier Flexible Retrieval: Archived data, slower access (minutes to hours), lower cost
7. S3 Glacier Deep Archive: Lowest cost, longest retrieval time (12-48 hours), rarely accessed data
8. S3 Outposts: On-premises object storage, local data residency, low-latency access

Amazon EBS (Elastic Block Store) vs Amazon S3 (Simple Storage Service)

Main differences:

Storage Type:
EBS: Block storage, which means data is stored in fixed-size blocks.
S3: Object storage, where each object is stored with a unique identifier.

Attachment/Access Method:
EBS: Attached directly to EC2 instances as a virtual hard drive.
S3: Accessed via API calls from anywhere on the internet.

Scalability:
EBS: Limited to the size of the volume you create (up to 16 TiB).
S3: Virtually unlimited storage capacity.

Pricing:
EBS: Charged based on the amount of provisioned storage and IOPS.
S3: Charged based on the amount of data stored, data transfer, and number of requests.

Example scenario:
Imagine you're running a photo-sharing application. You might use:
EBS for storing the application's operating system and database.
S3 for storing user-uploaded photos, which can be easily accessed and shared.

File storage

In file storage, multiple clients (such as users, applications, servers, and so on) 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 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 to petabytes without disrupting applications.

Amazon EBS (Elastic Block Store) vs Amazon EFS (Elastic File System)
EBS is for single-instance, high-performance block storage, while EFS is for multi-instance, scalable file storage.

Main differences:

Storage Type:
EBS: Block storage, suitable for applications that need low-latency access to data.

EFS: File storage, ideal for shared access across multiple instances.

Accessibility:
EBS: Can be attached to a single EC2 instance at a time within the same Availability Zone.
EFS: Can be accessed by multiple EC2 instances simultaneously across multiple Availability Zones in a region.
EBS volumes store data within a single Availability Zone. Amazon EFS file systems store data across multiple Availability Zones.

Scalability:
EBS: Fixed capacity that you specify (up to 16 TiB per volume).
EFS: Automatically scales up or down as you add or remove files, with no need to provision storage in advance.

Performance:
EBS: Provides consistent and low-latency performance.
EFS: Performance scales with the size of the file system.

Example scenario:
Imagine you're running a content management system:
You might use EBS for storing the database files, as they require low-latency access and are specific to a single EC2 instance.
You could use EFS for storing website assets (images, videos, etc.) that need to be accessed by multiple web server instances simultaneously.

Amazon Relational Database Service (Amazon RDS)

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. With these capabilities, you can spend less time completing administrative tasks and more time using data to innovate your applications. You can integrate Amazon RDS with other services to fulfill your business and operational needs, such as using AWS Lambda to query your database from a serverless application.

Amazon RDS provides a number of different security options. 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).

Amazon RDS database engines
Amazon RDS is available on six database engines, which optimize for memory, performance, or input/output (I/O). Supported database engines include:

- Amazon Aurora
- PostgreSQL
- MySQL
- MariaDB
- Oracle Database
- Microsoft SQL Server

Amazon Aurora

is an enterprise-class relational database. It is compatible with MySQL and PostgreSQL relational databases. It is up to five times faster than standard MySQL databases and up to three times faster than standard 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.

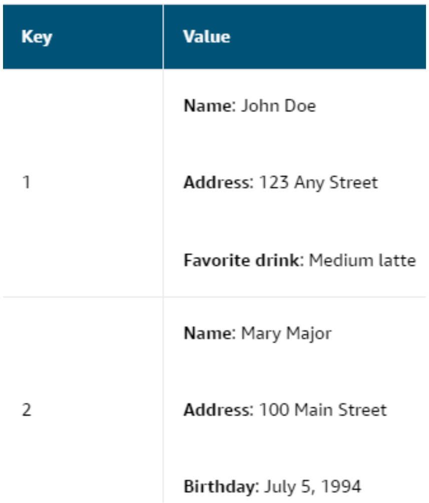
Consider Amazon Aurora if your workloads require high availability. It replicates six copies of your data across three Availability Zones and continuously backs up your data to Amazon S3.

Nonrelational databases

In a nonrelational database, you create tables. A table is a place 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 for nonrelational databases is key-value pairs. With key-value pairs, data is organized into items (keys), and items have attributes (values). You can think of attributes as being different features of your data.

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.



Key	Value
1	**Name:** John Doe **Address:** 123 Any Street **Favorite drink:** Medium latte
2	**Name:** Mary Major **Address:** 100 Main Street **Birthday:** July 5, 1994

Amazon DynamoDB

is a **key-value NoSQL** database service. It delivers **single-digit millisecond performance** at any scale. Support for both key-value and document data models

> DynamoDB is **serverless**, which means that you do not have to provision, patch, or manage servers. You also do not have to install, maintain, or operate software.

Auto Scaling:
As the size of your database shrinks or grows, DynamoDB automatically scales to adjust for changes in capacity while maintaining consistent performance. This makes it a suitable choice for use cases that require high performance while scaling.

Amazon RDS vs Amazon DynamoDB

1. Database Type:

RDS: Relational (SQL) databases with structured data and predefined schemas
DynamoDB: NoSQL database with flexible schemas for semi-structured data

2. Scalability:
RDS: Vertical scaling (increasing instance size) and some horizontal scaling with read replicas
DynamoDB: Seamless horizontal scaling with no downtime

3. Performance:
RDS: Performance depends on the instance type and storage
DynamoDB: Consistent single-digit millisecond latency regardless of scale

4. Pricing Model:
RDS: Charged based on instance hours, storage, and data transfer

DynamoDB: Charged based on read/write capacity units or on-demand pricing

5. Data Structure:
RDS: Tables with fixed schemas, supports complex joins
DynamoDB: Tables with flexible schemas, limited join capabilities

Example scenario:
Imagine you're building two different applications:

An e-commerce platform that requires complex queries and transactions:
Use Amazon RDS (e.g., MySQL) for structured data and support for complex joins.

A mobile game that needs to handle millions of users with low-latency data access:
Use Amazon DynamoDB for its ability to scale seamlessly and provide consistent performance.

Amazon Redshift

is 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.

- Redshift is a columnar storage database optimized for complex queries on large datasets.
- It's designed for online analytical processing (OLAP) and business intelligence applications.
- SQL Interface: Compatible with existing business intelligence tools.
- Like Snowflake or Google BigQuery

AWS Database Migration Service (AWS DMS)

enables you to migrate relational databases, nonrelational databases, and other types of data stores. Supports homogeneous and heterogeneous migrations.
For example, suppose that you have a MySQL database that is stored on premises in an Amazon EC2 instance or in Amazon RDS. Consider the MySQL database to be your source database. Using AWS DMS, you could migrate your data to a target database, such as an Amazon Aurora database.

1. Development and test database migrations:

Purpose: Create copies of production databases for development or testing environments. Enabling developers to test applications against production data without affecting production users

How DMS helps:
Allows creating a replica of the source database without affecting production.
Supports heterogeneous migrations (e.g., Oracle to MySQL).
Minimizes downtime during migration process.

2. Database consolidation:
Purpose: Combine multiple databases into a single target database.
How DMS helps:
Supports migrations from various source databases to a single AWS target.
Can consolidate on-premises databases to AWS-managed databases like RDS or Aurora.
Reduces operational costs and simplifies database management.

3. Continuous replication:
Purpose: Maintain synchronization between source and target databases.
How DMS helps:
Supports ongoing replication with minimal latency.
Useful for disaster recovery or distributing data across different regions.
Can be used to keep on-premises and cloud databases in sync during migration.

Additional database services

1. **Amazon DocumentDB** is a document database service that supports MongoDB workloads. (MongoDB is a document database program.)

2. **Amazon Neptune** is a graph database service. to build and run applications that work with highly connected datasets, such as recommendation engines, fraud detection, and knowledge graphs.

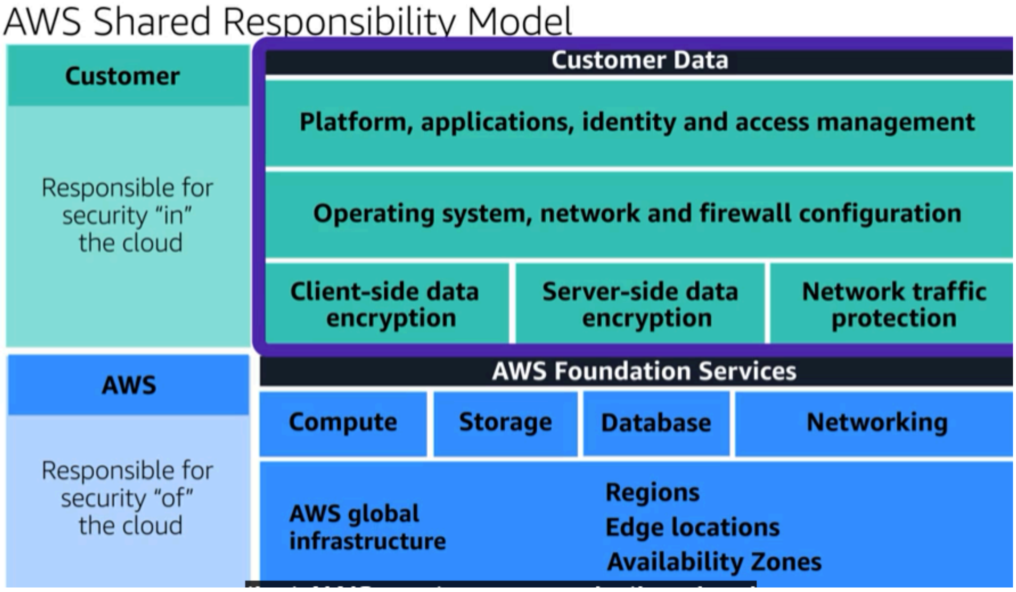
3. **Amazon Quantum Ledger Database (Amazon QLDB)** is a ledger database service. to review a complete history of all the changes that have been made to your application data.

4. **Amazon Managed Blockchain** is 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.

5. **Amazon ElastiCache** is a service that adds caching layers on top of your databases to help improve the read times of common requests. It supports two types of data stores: Redis and Memcached.

6. **Amazon DynamoDB Accelerator (DAX)** is an in-memory cache for DynamoDB. It helps improve response times from single-digit milliseconds to microseconds.

Module 6: Security



AWS Shared Responsibility Model

Customer

Responsible for security "in" the cloud

Customer Data

Platform, applications, identity and access management

Operating system, network and firewall configuration

Client-side data encryption

Server-side data encryption

Network traffic protection

AWS

Responsible for security "of" the cloud

AWS Foundation Services

Compute

Storage

Database

Networking

AWS global infrastructure

Regions

Edge locations

Availability Zones

The AWS shared responsibility model

The shared responsibility model divides into customer responsibilities (commonly referred to as "security in the cloud") and AWS responsibilities (commonly referred to as "security of the cloud").

<table>
<tr><td rowspan="4">Customers</td><td colspan="3">Customer Data</td></tr>
<tr><td colspan="3">Platform, Applications, Identity and Access Management</td></tr>
<tr><td colspan="3">Operating Systems, Network and Firewall Configuration</td></tr>
<tr><td>Client-side Data Encryption</td><td>Server-side Encryption</td><td>Networking Traffic Protection</td></tr>
</table>

<table>
<tr><td rowspan="4">AWS</td><td colspan="12">Software</td></tr>
<tr><td colspan="3">Compute</td><td colspan="3">Storage</td><td colspan="3">Database</td><td colspan="3">Networking</td></tr>
<tr><td colspan="12">Hardware/AWS Global Infrastructure</td></tr>
<tr><td colspan="4">Regions</td><td colspan="4">Availability Zones</td><td colspan="4">Edge Locations</td></tr>
</table>

Customers: Security in the cloud

Customers are responsible for the security of everything that they create and put in the AWS Cloud. Steps include selecting, configuring, and patching the operating systems that will run on Amazon EC2 instances, configuring security groups, and managing user accounts.

AWS: Security of the cloud
AWS is responsible for 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 manages the security of the cloud, specifically the physical infrastructure that hosts your resources, which include:

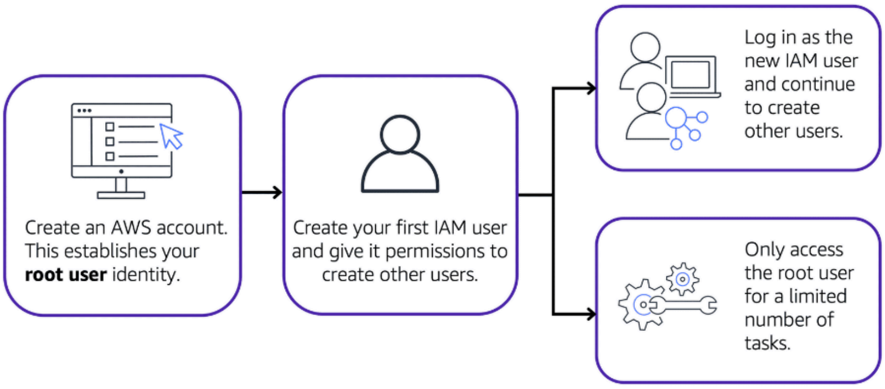
- Physical security of data centers
- Hardware and software infrastructure
- Network infrastructure
- Virtualization infrastructure

User Permissions and Access

AWS Identity and Access Management (IAM)

- IAM users, groups, and roles
- IAM policies
- Multi-factor authentication

AWS account root user
When you first create an AWS account, you begin with an identity known as the root user.


Create an AWS account. This establishes your **root user** identity.

Create your first IAM user and give it permissions to create other users.

Log in as the new IAM user and continue to create other users.

Only access the root user for a limited number of tasks.

Best practice:
Do not use the root user for everyday tasks.
Instead, use the root user to create your first IAM user and assign it permissions to create other users.

Then, continue to create other IAM users, and access those identities for performing regular tasks throughout AWS. Only use the root user when you need to perform a limited number of tasks that are only available to the root user. Examples of these tasks include changing your root user email address and changing your AWS support plan.

IAM users

An IAM user is an identity that you create in AWS. It represents the person or application that interacts with AWS services and resources. 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.

Best practice:

We recommend that you create individual IAM users for each person who needs to access AWS.

Even if you have multiple employees who require the same level of access, you should create individual IAM users for each of them. This provides additional security by allowing each IAM user to have a unique set of security credentials.

IAM policies

An IAM policy is a *document* that allows/grants or denies permissions to AWS services and resources.

IAM policies enable you to customize users' levels of access to resources. For example, you can allow users to access all of the Amazon S3 buckets within your AWS account, or only a specific bucket.

Best practice:

Follow the security principle of *least privilege* when granting permissions.

By following this principle, you help to prevent users or roles from having more permissions than needed to perform their tasks.

For example, if an employee needs access to only a specific bucket, specify the bucket in the IAM policy. Do this instead of granting the employee access to all of the buckets in your AWS account.

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.

Assigning IAM policies at the group level also makes it easier to adjust permissions when an employee transfers to a different job. For example, if a cashier becomes an inventory specialist, the coffee shop owner removes them from the "Cashiers" IAM group and adds them into the "Inventory Specialists" IAM group.

IAM roles

When the employee needs to switch to a different task, they give up their access to one workstation and gain access to the next workstation. The employee can easily switch between workstations, but at any given point in

time, they can have access to only a single workstation. This same concept exists in AWS with IAM roles.
An IAM role is an identity that you can assume to gain temporary access to permissions.

Best practice:
IAM roles are ideal for situations in which access to services or resources needs to be granted *temporarily*, instead of long-term.

Multi-factor authentication (MFA)
authentication response from AWS MFA device. This device could be a hardware security key, a hardware device, or an MFA application on a device such as a smartphone.
can enable MFA for the root user and IAM users. As a best practice, enable MFA for the root user and all IAM users in your account.

AWS Organizations

Suppose that your company has multiple AWS accounts. You can use AWS Organizations to consolidate and manage multiple AWS accounts within a central location.

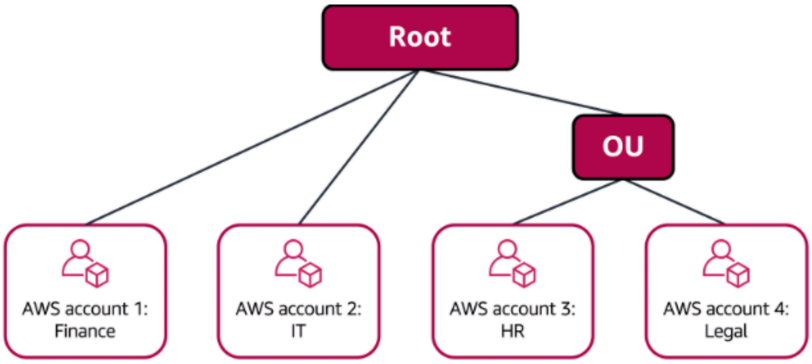
When you create an organization, AWS Organizations automatically creates a root, which is the parent container for all the accounts in your organization.

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.
Consolidated billing is another feature of AWS Organizations.

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.

By organizing separate accounts into OUs, you can more easily isolate workloads or applications that have specific security requirements. For instance, if your company has accounts that can access only the AWS services that meet certain regulatory requirements, you can put these accounts into one OU. Then, you can attach a policy to the OU that blocks access to all other AWS services that do not meet the regulatory requirements.



The HR and legal departments need to access the same AWS services and resources, so you place them into an OU together. Placing them into an OU empowers you to attach policies that apply to both the HR and legal departments' AWS accounts.

In AWS Organizations, you can apply service control policies (SCPs) to the organization root, an individual member account, or an OU. An SCP affects all IAM users, groups, and roles within an account, including the AWS account root user.
You can apply IAM policies to IAM users, groups, or roles. You cannot apply an IAM policy to the AWS account root user.

Compliance

AWS Artifact

Depending on your company's industry, you may need to uphold specific standards. An audit or inspection will ensure that the company has met those standards.

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: AWS Artifact Agreements and AWS Artifact Reports.

AWS Artifact Agreements

Suppose that your company needs to *sign an agreement* with AWS regarding your use of certain types of information throughout AWS services. You can do this through AWS Artifact Agreements.

In AWS Artifact 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 to address the needs of customers who are subject to specific regulations, such as the Health Insurance Portability and Accountability Act (HIPAA).

AWS Artifact Reports

Next, suppose that a member of your company's development team is building an application and needs more information about their responsibility for complying with certain regulatory standards. You can advise them to access this information in AWS Artifact Reports.

AWS Artifact 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. AWS Artifact Reports remains up to date with the latest reports released. You can provide the AWS audit artifacts to your auditors or regulators as evidence of AWS security controls.

AWS Artifact provides access to AWS security and compliance documents, such as AWS ISO certifications, Payment Card Industry (PCI) reports, and Service Organization Control (SOC) reports.

Customer Compliance Center

in the Customer Compliance Center, you can read customer compliance stories to discover how companies in regulated industries have solved various compliance, governance, and audit challenges.

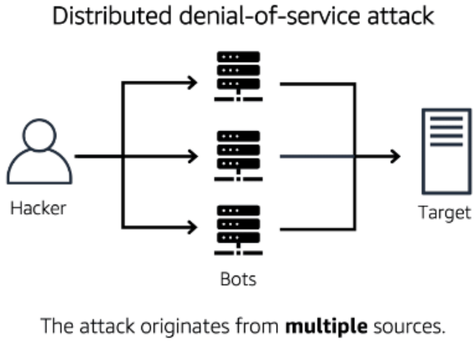
You can also access compliance whitepapers and documentation on topics such as:

AWS answers to key compliance questions
An overview of AWS risk and compliance
An auditing security checklist
Additionally, the Customer Compliance Center includes an auditor learning path.

Denial-of-Service Attacks
A denial-of-service (DoS) attack is a deliberate attempt to make a website or application unavailable to users.

Distributed denial-of-service attacks DDoS.

Now, suppose that the prankster has enlisted the help of friends. multiple sources are used to start an attack. Group of attackers or The single attacker can use multiple infected computers (also known as “bots”) to send excessive traffic to a website or application. To help minimize the effect of DoS and DDoS attacks on your applications, can use AWS Shield


Distributed denial-of-service attack
Hacker
Bots
Target
The attack originates from **multiple** sources.

AWS Shield
AWS Shield is a service that protects applications against DDoS attacks. AWS Shield provides two levels of protection: Standard and Advanced.

AWS Shield 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, AWS Shield Standard uses a variety of analysis techniques to detect malicious traffic in real time and automatically mitigates it.

AWS Shield Advanced is a paid service that provides detailed attack diagnostics and the ability to detect and mitigate sophisticated DDoS attacks.
It also integrates with other services such as Amazon CloudFront, Amazon Route 53, and Elastic Load Balancing. Additionally, you can integrate AWS Shield with AWS WAF by writing custom rules to mitigate complex DDoS attacks.

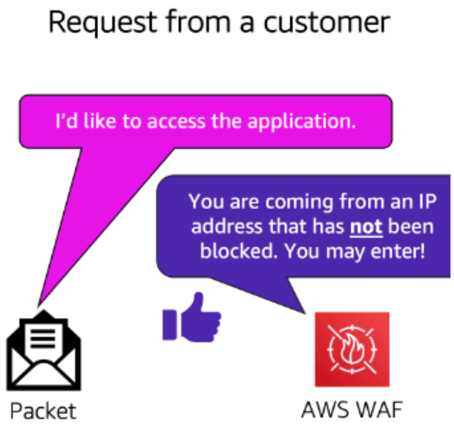
AWS Security Services

AWS Key Management Service (AWS KMS)

To ensure that your applications' data is secure while in storage (encryption at rest) and while it is transmitted, known as encryption in transit, AWS KMS enables you to perform encryption operations through the use of cryptographic keys. A cryptographic key is a random string of digits used for locking (encrypting) and unlocking (decrypting) data. You can use AWS KMS to create, manage, and use cryptographic keys. 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 WAF

AWS WAF is a web application firewall that lets you monitor network requests that come into your web applications.
WAF works together with Amazon CloudFront and an Application Load Balancer. Recall the network access control, WAF works in a similar way to block or allow traffic. However, it does this by using a web access control list (ACL) to protect your AWS resources.


Request from a customer
I'd like to access the application.
You are coming from an IP address that has not been blocked. You may enter!
Packet
AWS WAF

When a request comes into AWS WAF, it checks against the list of rules that you have configured in the web ACL. If a request does not come from one of the blocked IP addresses, it allows access to the application. However, if a request comes from one of the blocked IP addresses that you have specified in the web ACL, AWS WAF denies access.

Amazon Inspector

perform automated security assessments.
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, such as open access to Amazon EC2 instances and installations of vulnerable software versions.
After Amazon Inspector has performed an assessment, it provides you with a list of security findings. The list prioritizes by severity level, including a detailed description of each security issue and a recommendation for how to fix it. Under the shared responsibility model, customers are responsible for the security of their applications, processes, and tools that run on AWS services.

Amazon GuardDuty

Amazon GuardDuty 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.

GuardDuty begins monitoring your network and account activity. You do not have to deploy or manage any additional security software. GuardDuty then continuously analyzes data from multiple AWS sources, including VPC Flow Logs and DNS logs.
If GuardDuty detects any threats, you can review detailed findings about them from the AWS Management Console. Findings include recommended steps for remediation. You can also configure AWS Lambda functions to take remediation steps automatically in response to GuardDuty's security findings.

Service Name	Focus	Scope	Operation
IAM	**Access** control and **user** management	AWS account-wide	Continuous, policy-based
Amazon Cognito	**User authentication and authorization for applications**	Application-level	On-demand, integrated with apps
AWS WAF	Web application protection from common exploits	Application layer (Layer 7)	Rule-based, continuous monitoring
Amazon GuardDuty	Intelligent **threat detection**	Account-wide, including VPC flow logs, CloudTrail, and DNS logs	Continuous, **machine learning-based**
Amazon Inspector	**Automated security assessments**	EC2 instances and container images	On-demand or scheduled scans
AWS Artifact	Compliance documentation access	**AWS compliance reports and agreements**	On-demand, manual access
AWS Shield	DDoS protection	Network and transport layers (Layer 3 and 4)	Automatic (Standard) or managed (Advanced)
AWS KMS	Encryption key management	AWS services and customer applications	On-demand key generation and management
Amazon **Macie**	**Data security and privacy**	**S3 buckets and objects**	**Continuous, machine learning-based discovery and protection**
AWS CloudHSM	Hardware-based key storage	Dedicated Hardware Security Modules (HSMs)	On-demand, customer-managed

AWS Secrets Manager	Secrets management (e.g., database credentials, API keys)	Application and service secrets	On-demand retrieval and rotation
AWS Security Hub	Centralized security management	Multi-account, integrates with other security services	Continuous monitoring and compliance checks

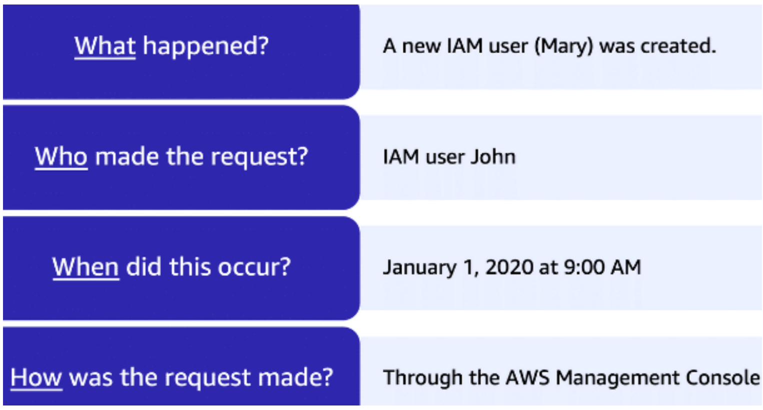
Module 7: Monitoring & Analytics

Amazon CloudWatch

a web service that enables you to monitor and manage various metrics and configure alarm actions based on data from those metrics.
CloudWatch uses metrics to represent the data points for your resources. AWS services send metrics to CloudWatch. CloudWatch then uses these metrics to create graphs automatically that show how performance has changed over time.
With CloudWatch, you can create alarms that automatically perform actions if the value of your metric has gone above or below a predefined threshold. you could create a CloudWatch alarm that automatically stops an Amazon EC2 instance when the CPU utilization percentage has remained below a certain threshold for a specified period. When configuring the alarm, you can specify to receive a notification whenever this alarm is triggered. For example, you can use a CloudWatch dashboard to monitor the CPU utilization of an Amazon EC2 instance, the total number of requests made to an Amazon S3 bucket, and more.

AWS CloudTrail

records API calls for your account. 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. You can think of CloudTrail as a "trail" of breadcrumbs (or a log of actions) that someone has left behind them.
Events are typically updated in CloudTrail within 15 minutes after an API call. You can filter events by specifying the time and date that an API call occurred, the user who requested the action, the type of resource that was involved in the API call, and more.



What happened?	A new IAM user (Mary) was created.
Who made the request?	IAM user John
When did this occur?	January 1, 2020 at 9:00 AM
How was the request made?	Through the AWS Management Console

Within CloudTrail, you can also enable CloudTrail Insights. ***This optional feature allows CloudTrail to automatically detect unusual API activities in your AWS account.***

Feature	Amazon CloudWatch	AWS CloudTrail
Primary Purpose	Monitoring and observability	Auditing and compliance
Type of Data	Metrics, logs, and events	API activity and account events
Main Use Cases	- Performance monitoring- Resource utilization tracking- Application and infrastructure alerting	- Security analysis- Compliance auditing- Operational troubleshooting
Data Collection	- Collects data from AWS services, applications, and on-premises resources- Can collect custom metrics	- Records AWS account activity- Logs all API calls made to AWS services
Retention Period	Configurable, typically 15 months for metrics	90 days by default, can be extended indefinitely using S3
Alerting Capabilities	Supports alarms based on metric thresholds	Can trigger CloudWatch Events (now EventBridge) for specific API calls
Integration	Integrates with many AWS services for monitoring	Integrates with compliance and security services like AWS Config

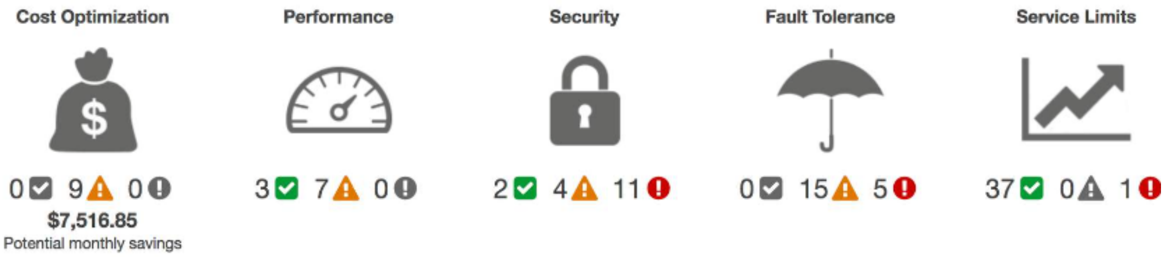
AWS Trusted Advisor

a web service that inspects your AWS environment and provides real-time recommendations in accordance with AWS best practices. The inspection/reviews 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,
- 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.

The guidance provided by AWS Trusted Advisor can benefit your company at all stages of deployment. For example, you can use AWS Trusted Advisor to assist you while you are creating new workflows and developing new applications. You can also use it while you are making ongoing improvements to existing applications and resources.



1. CloudWatch is your go-to for ongoing monitoring, performance tracking, and operational alerts.
2. CloudTrail is essential for security, audit, and compliance, tracking who did what in your AWS account.
3. Trusted Advisor provides holistic recommendations across multiple categories, helping you follow AWS best practices.

Module 8: Pricing & Support

AWS Free Tier

Three types of offers are available:

- Always Free:
  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:
  Examples include specific amounts of Amazon S3 Standard Storage, thresholds for monthly hours of Amazon EC2 compute time, and amounts of Amazon CloudFront data transfer out.
- 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. For example, Amazon Inspector offers a 90-day free trial. Amazon Lightsail (a service that enables you to run virtual private servers) offers 750 free hours of usage over a 30-day period.

How AWS pricing works

AWS offers a range of cloud computing services with pay-as-you-go pricing.

The AWS Pricing Calculator 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.
In the AWS Pricing Calculator, you can enter details, such as the kind of operating system you need, memory requirements, and input/output (I/O) requirements. By using the AWS Pricing Calculator, you can review an estimated comparison of different EC2 instance types across AWS Regions.

Examples for Lambda & EC2 Pricings:

For **AWS Lambda**, you are charged based on the number of requests for your functions and the time that it takes for them to run.
AWS Lambda allows 1 million free requests and up to 3.2 million seconds of compute time per month.
You can save on AWS Lambda costs by signing up for a Compute Savings Plan. A Compute Savings Plan offers lower compute costs in exchange for committing to a consistent amount of usage over a 1-year or 3-year term. This is an example of paying less when you reserve. If you have used AWS Lambda in multiple AWS Regions, you can view the itemized charges by Region on your bill.

With **Amazon EC2**, you pay for only the compute time that you use while your instances are running.
For some workloads, you can significantly reduce Amazon EC2 costs by using Spot Instances. For example, suppose that you are running a batch processing job that is able to withstand interruptions. Using a Spot Instance would provide you with up to 90% cost savings while still meeting the availability requirements of your workload. Savings Plans and Reserved Instances save money.

For **Amazon S3** pricing, consider the following cost components:

- Storage - You pay for only the storage that you use. You are charged the rate to store objects in your Amazon S3 buckets based on your objects' sizes, storage classes, and how long you have stored each object during the month.
- Requests and data retrievals - You pay for requests made to your Amazon S3 objects and buckets. For example, suppose that you are storing photo files in Amazon S3 buckets and hosting them on a website. Every time a visitor requests the website that includes these photo files, this counts towards requests you must pay for.
- 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. However, you pay for data that you transfer into and out of Amazon S3, with a few exceptions. There is no cost for data transferred into Amazon S3 from the internet or out to Amazon CloudFront. There is also no cost for data transferred out to an Amazon EC2 instance in the same AWS Region as the Amazon S3 bucket.
- 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.

Billing Dashboard

Use the AWS Billing & Cost Management dashboard(opens in a new tab) to 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.

Another benefit of consolidated billing is the ability to share bulk discount pricing, Savings Plans, and Reserved Instances across the accounts in your organization. For instance, one account might not have enough monthly usage to qualify for discount pricing. However, when multiple accounts are combined, their aggregated usage may result in a benefit that applies across all accounts in the organization.

Consolidated billing also enables you to share volume pricing discounts across accounts.

Some AWS services, such as Amazon S3, provide volume pricing discounts that give you lower prices the more that you use the service. In Amazon S3, after customers have transferred 10 TB of data in a month, they pay a lower per-GB transfer price for the next 40 TB of data transferred.

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 budgeted amounts or to the AWS Free Tier limits.

In AWS Budgets, you can also set custom alerts when your usage exceeds (or is forecasted to exceed) the budgeted amount.

AWS Cost Explorer

a tool that lets you visualize, understand, and manage your AWS costs and usage over time.

AWS Cost Explorer includes a default report of the costs and usage for your top five cost-accruing AWS services. You can apply custom filters and groups to analyze your data. For example, you can view resource usage at the hourly level.

AWS Support

AWS offers 4 different Support plans to help you troubleshoot issues, lower costs, and efficiently use AWS services.

You can choose from the following Support plans to meet your company's needs:

- Basic
- Developer
- Business
- Enterprise On-Ramp
- Enterprise

1. **Basic Support** is free for all AWS customers. It includes access to whitepapers, documentation, and support communities. With Basic Support, 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.

2. Customers in the **Developer Support plan** have access to features such as:

- Best practice guidance
- Client-side diagnostic tools
- Building-block architecture support, which consists of guidance for how to use AWS offerings, features, and services together

3. Customers with a **Business Support plan** have access to additional features, including:

- Use-case guidance to identify AWS offerings, features, and services that can best support your specific needs
- All AWS Trusted Advisor checks
- Limited support for third-party software, such as common operating systems and application stack components

4. In November 2021, AWS opened enrollment into **AWS Enterprise On-Ramp Support** plan. In addition to all the features included in the Basic, Developer, and Business Support plans, customers with an Enterprise On-Ramp Support plan have access to:

- A pool of Technical Account Managers to provide proactive guidance and coordinate access to programs and AWS experts
- A Cost Optimization workshop (one per year)
- A Concierge support team for billing and account assistance
- Tools to monitor costs and performance through Trusted Advisor and Health API/Dashboard

Enterprise On-Ramp Support plan also provides access to a specific set of proactive support services, which are provided by a pool of Technical Account Managers.

- Consultative review and architecture guidance (one per year)
- Infrastructure Event Management support (one per year)
- Support automation workflows
- 30 minutes or less response time for business-critical issues

5. customers with **Enterprise Support** have access to:

- A designated Technical Account Manager to provide proactive guidance and coordinate access to programs and AWS experts
- A Concierge support team for billing and account assistance
- Operations Reviews and tools to monitor health
- Training and Game Days to drive innovation
- Tools to monitor costs and performance through Trusted Advisor and Health API/Dashboard

The Enterprise plan also provides full access to proactive services, which are provided by a designated Technical Account Manager:

- Consultative review and architecture guidance
- Infrastructure Event Management support
- Cost Optimization Workshop and tools
- Support automation workflows
- 15 minutes or less response time for business-critical issues

Technical Account Manager (TAM)
The Enterprise On-Ramp and Enterprise Support plans include access to a Technical Account Manager (TAM).
The TAM is your primary point of contact at AWS. If your company subscribes to Enterprise Support or Enterprise On-Ramp, your TAM educates, empowers, and evolves your cloud journey across the full range of AWS services. TAMs provide expert engineering guidance, help you design solutions that efficiently integrate AWS services, assist with cost-effective and resilient architectures, and provide direct access to AWS programs and a broad community of experts.

comparison of AWS Support plans, focusing on key differentiators:

Basic Support:

- Available to all AWS customers
- Access to AWS Trusted Advisor core checks
- AWS Health Dashboard

Developer Support:

- Response times: General guidance < 24 hours, System impaired < 12 hours
- Business hours access to Cloud Support Associates
- General architectural guidance

Business Support

- Faster response times: Production system down < 1 hour
- 24/7 phone, web, and chat access to Cloud Support Engineers
- Full set of Trusted Advisor checks
- Use-case specific architectural guidance
- Third-party software support
- AWS Support API access

Enterprise On-Ramp Support:

- Even faster response times: Business-critical system down < 30 minutes
- Pool of Technical Account Managers (TAMs)
- Annual consultative review
- Access to proactive services (e.g., one Infrastructure Event Management per year)
- Support Automation Workflows

Enterprise Support:

- Fastest response times: Business/Mission-critical system down < 15 minutes
- Designated Technical Account Manager (TAM)
- Consultative reviews based on applications
- Full access to proactive services
- AWS Incident Detection and Response (additional fee)

Key exam-relevant points:

1. Response times improve as you move up support tiers
2. Only Business and above offer full Trusted Advisor checks
3. TAM access starts at Enterprise On-Ramp (pooled) and becomes designated in Enterprise
4. Third-party software support begins with Business plan
5. Proactive services increase with higher-tier plans

For the exam, you might encounter scenario-based questions like:
"A company needs architectural guidance specific to their use case, full Trusted Advisor checks, and 24/7 support. Which is the minimum AWS Support plan that meets these requirements?"
The correct answer would be the Business Support plan, as it's the lowest tier that offers all these features.

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.

Module 9: Migration & Innovation

AWS Cloud Adoption Framework (AWS CAF)

At the highest level, the AWS Cloud Adoption Framework (AWS CAF) organizes guidance into six areas of focus, called **Perspectives**. Each Perspective addresses distinct responsibilities.
In general, the Business, People, and Governance Perspectives focus on business capabilities, whereas the Platform, Security, and Operations Perspectives focus on technical capabilities.

1. **Business Perspective:**

ensures that IT aligns with business needs and that IT investments link to key business results.
Use the Business Perspective to create a strong business case for cloud adoption and prioritize cloud adoption initiatives. Ensure that your business strategies and goals align with your IT strategies and goals.
Common roles in the Business Perspective include:

- Business managers
- Finance managers
- Budget owners
- Strategy stakeholders

2. People Perspective

supports development of an organization-wide change management strategy for successful cloud adoption.
Use the People Perspective to evaluate organizational structures and roles, new skill and process requirements, and identify gaps. This helps prioritize training, staffing, and organizational changes.
Common roles in the People Perspective include:

- Human resources
- Staffing
- People managers

3. Governance Perspective

focuses on the skills and processes to align IT strategy with business strategy. This ensures that you maximize the business value and minimize risks.
Use the Governance Perspective to understand how to update the staff skills and processes necessary to ensure business governance in the cloud. Manage and measure cloud investments to evaluate business outcomes.
Common roles in the Governance Perspective include:

- Chief Information Officer (CIO)
- Program managers
- Enterprise architects
- Business analysts
- Portfolio managers

4. Platform Perspective

includes principles and patterns for implementing new solutions on the cloud, and migrating on-premises workloads to the cloud.

Use a variety of architectural models to understand and communicate the structure of IT systems and their relationships. Describe the architecture of the target state environment in detail.

Common roles in the Platform Perspective include:

- Chief Technology Officer (CTO)
- IT managers
- Solutions architects

5. Security Perspective

ensures that the organization meets security objectives for visibility, auditability, control, and agility.

Use the AWS CAF to structure the selection and implementation of security controls that meet the organization's needs.

Common roles in the Security Perspective include:

- Chief Information Security Officer (CISO)
- IT security managers
- IT security analysts

6. Operations Perspective

helps you to enable, run, use, operate, and recover IT workloads to the level agreed upon with your business stakeholders.

Define how day-to-day, quarter-to-quarter, and year-to-year business is conducted. Align with and support the operations of the business. The AWS CAF helps these stakeholders define current operating procedures and identify the process changes and training needed to implement successful cloud adoption.

Common roles in the Operations Perspective include:

- IT operations managers
- IT support managers

6 R's strategies for migration

When migrating applications to the cloud, six of the most common migration strategies that you can implement are:

1. **Rehosting**
2. **Replatforming**
3. **Refactoring/re-architecting**
4. **Repurchasing**
5. **Retaining**
6. **Retiring**

Rehosting also known as “lift-and-shift” involves moving applications without changes.
In the scenario of a large legacy migration, in which the company is looking to implement its migration and scale quickly to meet a business case, the majority of applications are rehosted.

Replatforming, also known as “lift, tinker, and shift,” involves making a few cloud optimizations to realize a tangible benefit. Optimization is achieved without changing the core architecture of the application.

Refactoring (also known as re-architecting) involves reimagining how an application is architected and developed by using cloud-native features. Refactoring is driven by a strong business need to add features, scale, or performance that would otherwise be difficult to achieve in the application’s existing environment.

Repurchasing involves moving from a traditional license to a software-as-a-service model.
For example, a business might choose to implement the repurchasing strategy by migrating from a customer relationship management (CRM) system to Salesforce.com.

Retaining consists of keeping applications that are critical for the business in the source environment. This might include applications that require major refactoring before they can be migrated, or, work that can be postponed until a later time.

Retiring is the process of removing applications that are no longer needed.

AWS Snow Family

The AWS Snow Family is a collection of physical devices that help to physically transport up to exabytes of data into and out of AWS.

AWS Snow Family is composed of AWS Snowcone, AWS Snowball, and AWS Snowmobile.
These devices offer different capacity points, and most include built-in computing capabilities. AWS owns and manages the Snow Family devices and integrates with AWS security, monitoring, storage management, and computing capabilities.

1. AWS Snowcone is a small, rugged, and secure edge computing and data transfer device.

It features 2 CPUs, 4 GB of memory, and up to 14 TB of usable storage.

2. AWS Snowball offers two types of devices:

a) Snowball Edge Storage Optimized devices are well suited for large-scale data migrations and recurring transfer workflows, in addition to local computing with higher capacity needs.

Storage: *80 TB of hard disk drive (HDD) capacity for block volumes and Amazon S3 compatible object storage, and 1 TB of SATA solid state drive (SSD) for block volumes.*

Compute: 40 vCPUs, and 80 GiB of memory to support Amazon EC2 sbe1 instances (equivalent to C5).

b) Snowball Edge Compute Optimized provides powerful computing resources for use cases such as machine learning, full motion video analysis, analytics, and local computing stacks.

Storage: *80-TB usable HDD capacity for Amazon S3 compatible object storage or Amazon EBS compatible block volumes and 28 TB of usable NVMe SSD capacity for Amazon EBS compatible block volumes.*

Compute: 104 vCPUs, 416 GiB of memory, and an optional NVIDIA Tesla V100 GPU. Devices run Amazon EC2 sbe-c and sbe-g instances, which are equivalent to C5, M5a, G3, and P3 instances.

3. AWS Snowmobile(opens in a new tab) is an exabyte-scale data transfer service used to move large amounts of data to AWS.

You can transfer up to *100 petabytes* of data per Snowmobile, a 45-foot long ruggedized shipping container, pulled by a semi trailer truck.

Innovate with AWS Services

When examining how to use AWS services, it is important to focus on the desired outcomes. You are properly equipped to drive innovation in the cloud if you can clearly articulate the following conditions:

- The current state
- The desired state
- The problems you are trying to solve

Serverless Applications

With AWS, serverless 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. AWS handles these capabilities for you.

AWS Lambda is an example of a service that you can use to run serverless applications. If you design your architecture to trigger Lambda functions to run your code, you can bypass the need to manage a fleet of servers.

Building your architecture with serverless applications enables your developers to focus on their core product instead of managing and operating servers.

AI

AWS offers a variety of services powered by artificial intelligence (AI).

For example, you can perform the following tasks:

- Convert speech to text with Amazon Transcribe.
- Discover patterns in text with Amazon Comprehend.
- Identify potentially fraudulent online activities with Amazon Fraud Detector.
- Build voice and text chatbots with Amazon Lex.

ML
Traditional machine learning (ML) development is complex, expensive, time consuming, and error prone. AWS offers Amazon SageMaker to remove the difficult work from the process and empower you to build, train, and deploy ML models quickly.
You can use ML to analyze data, solve complex problems, and predict outcomes before they happen.

Amazon Q Developer
Amazon Q Developer is a machine learning-powered code generator that provides you with code recommendations in real time. You can also activate Amazon Q Developer from directly within the AWS Lambda and AWS Cloud9 console code editors.

Migration Services Review

AWS DMS can be used to migrate data from an on-premises database to a database in AWS. However, AWS DMS does not migrate the actual server to an EC2 instance.

Migration Hub is a service that helps plan and track application migrations. Migration Hub does not perform system migrations.

AWS MGN is an automated lift-and-shift solution. This solution can migrate physical servers and any databases or applications that run on them to EC2 instances in AWS.

Application Discovery Service collects information about the usage and configuration of on-premises servers to help plan a migration to AWS. Application Discovery Service does not actually perform migration operations.

Module 10: The Cloud Journey

Well-Architected Framework

WAF 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.

6 Pillars

1. Operational excellence

is the ability to run and monitor systems to deliver business value and to continually improve supporting processes and procedures.
Design principles for operational excellence in the cloud include performing operations as code, annotating documentation, anticipating failure, and frequently making small, reversible changes.

2. Security

is the ability to protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies.
When considering the security of your architecture, apply these best practices:

- Automate security best practices when possible.
- Apply security at all layers.
- Protect data in transit and at rest.

3. Reliability

is the ability of a system to do the following:

- Recover from infrastructure or service disruptions
- Dynamically acquire computing resources to meet demand
- Mitigate disruptions such as misconfigurations or transient network issues

Reliability includes testing recovery procedures, scaling horizontally to increase aggregate system availability, and automatically recovering from failure.

❖ focuses on the ability of a workload to consistently and correctly perform its intended functions

4. Performance efficiency

is the ability to use computing resources efficiently to meet system requirements and to maintain that efficiency as demand changes and technologies evolve.

Evaluating the performance efficiency of your architecture includes experimenting more often, using serverless architectures, and designing systems to be able to go global in minutes.

5. Cost optimization

is the ability to run systems to deliver business value at the lowest price point.
Cost optimization includes adopting a consumption model, analyzing and attributing expenditure, and using managed services to reduce the cost of ownership.

6. Sustainability

is the ability to continually improve sustainability impacts by reducing energy consumption and increasing efficiency across all components of a workload by maximizing the benefits from the provisioned resources and minimizing the total resources required.
To facilitate good design for sustainability:

- Understand your impact
- Establish sustainability goals
- Maximize utilization
- Anticipate and adopt new, more efficient hardware and software offerings
- Use managed services
- Reduce the downstream impact of your cloud workloads

Review

Pillar	Focus	Key Concepts	Remember As
Operational Excellence	Running and monitoring systems	- Automation- Continuous improvement- Observability	"Smooth Operations"
Security	Protecting data and systems	- Identity management- Encryption- Threat detection	"Protect Everything"
Reliability	System recovery and availability	- Fault tolerance- Disaster recovery- Scalability	"Always Works"
Performance Efficiency	Using resources efficiently	- Right sizing- Monitoring- Optimizing	"Fast and Lean"
Cost Optimization	Avoiding unnecessary costs	- Resource allocation- Matching supply and demand- Expenditure awareness	"Smart Spending"
Sustainability	Minimizing environmental impact	- Energy efficiency- Resource optimization- Sustainable practices	"Green Cloud"

1. Operational Excellence: Think "How smoothly does it run?"
2. Security: Ask "How well is it protected?"
3. Reliability: Consider "Will it work when needed?"
4. Performance Efficiency: Question "How fast and resource-efficient is it?"
5. Cost Optimization: Ponder "Are we spending wisely?"
6. Sustainability: Reflect "How environmentally friendly is it?"

Benefits of the AWS Cloud

6 advantages of cloud computing

1. Trade upfront expense for variable expense:

   Upfront expenses include data centers, physical servers, and other resources that you would need to invest in before using computing resources.

   Instead of investing heavily in data centers and servers before you know how you're going to use them, you can pay only when you consume computing resources.

2. Benefit from massive economies of scale:

   By using cloud computing, you can achieve a lower variable cost than you can get on your own.

   Because usage from hundreds of thousands of customers aggregates in the cloud, AWS can achieve higher economies of scale. **Economies of scale translate into lower pay-as-you-go prices.**

3. Stop guessing capacity: With cloud computing, you don't have to predict how much infrastructure capacity you will need before deploying an application. Instead of paying for resources that are unused or dealing with limited capacity, you can access only the capacity that you need, and scale in or out in response to demand.

4. Increase speed and agility:

   The flexibility of cloud computing makes it easier for you to develop and deploy applications.

   This flexibility also provides your development teams with more time to experiment and innovate.

5. Stop spending money running and maintaining data centers:

   Cloud computing in data centers often requires you to spend more money and time managing infrastructure and servers.

   A benefit of cloud computing is the ability to focus less on these tasks and more on your applications and customers.

6. Go global in minutes: The AWS Cloud global footprint enables you to quickly deploy applications to customers around the world, while providing them **with low latency.**

In Scope AWS Services & Features

Analytics

Amazon Athena

- Key points: Serverless, interactive query service for S3 data
- Use cases: Ad-hoc queries on log files, business reports
- Exam focus: Cost-effective solution for querying data in S3 without loading it into a database

AWS Data Exchange

- Key points: Find, subscribe to, and use third-party data in the cloud
- Use cases: Acquiring datasets for analysis, machine learning, or app development
- Exam focus: Easy way to find and use external data sources

Amazon EMR (Elastic MapReduce)

- Key points: Managed big data platform for processing vast amounts of data
- Use cases: Log analysis, financial analysis, genomics
- Exam focus: Scalable solution for big data processing and analysis

AWS Glue

- Key points: Fully managed extract, transform, and load (ETL) service
- Use cases: Preparing and loading data for analytics
- Exam focus: Serverless data integration service that makes data preparation easier

Amazon Kinesis

- Key points: Real-time **data streaming** and analysis
- Use cases: Log and event data collection, real-time analytics
- Exam focus: Ability to process and analyze real-time, streaming data

Service	Key Purpose	Use Cases	Exam Tips
Amazon Athena	Serverless query service for S3 data	- Ad-hoc queries on log files- Business reports	- No-setup querying of S3 data- Pay per query
AWS Data Exchange	Marketplace for third-party data	- Acquiring datasets for analysis- ML model training	- Finding external data sources- Subscribing to datasets
Amazon EMR	Managed big data processing	- Log analysis- Genomics- Financial analysis	- Works with Hadoop, Spark- Scalable cluster management
AWS Glue	Fully managed ETL service	- Data preparation for analytics- Data catalog creation	- Serverless data integration- Automated schema discovery
Amazon Kinesis	Real-time data streaming and analysis	- Log and event data collection- Real-time analytics	- Processing streaming data- Real-time insights
Amazon MSK	Managed Kafka service	- Building data pipelines- Streaming analytics	- Fully compatible with Apache Kafka- Serverless option available
Amazon OpenSearch Service	Search and analytics engine	- Log analytics- Full-text search	- Formerly Amazon Elasticsearch Service- Visualize with OpenSearch Dashboards
Amazon QuickSight	Business intelligence service	- Creating interactive dashboards- Embedding analytics in apps	- Pay-per-session pricing- ML-powered insights
Amazon Redshift	Data warehousing	- Business intelligence- Big data analytics	- Petabyte-scale data warehouse- Columnar storage for fast queries

Application Integration

services:

Service	Key Purpose	Use Cases	Exam Tips
Amazon EventBridge	Serverless event bus	- Event-driven architectures- Connecting applications	- Reacts to changes in AWS and custom apps- Replaces CloudWatch Events
Amazon SNS	Pub/sub messaging	- Sending alerts- Mobile push notifications	- Push messages to multiple subscribers- Supports various protocols (HTTP, email, SMS)
Amazon SQS	Message queuing	- Decoupling components- Handling high-volume messaging	- Asynchronous communication- Ensures reliable message delivery
AWS Step Functions	Workflow orchestration	- Coordinating multiple AWS services- Building complex processes	- Visual workflow creation- Manages state and error handling

Exam Tips: Understand the primary function of each service (e.g., EventBridge for events, SNS for notifications, SQS for queues, Step Functions for workflows).

Business Applications

services:

Service	Key Purpose	Use Cases	Exam Tips
Amazon Connect	Cloud-based contact center	- Customer service call centers- Support desks	- Easy to set up and scale- Pay-as-you-go pricing- Integrates with other AWS services
Amazon SES	Email sending service	- Transactional emails- Marketing communications	- High deliverability rates- Pay only for what you use- Can be used with other AWS services

Exam Tips:
Understand the primary function of each service (Connect for contact centers, SES for email).

Financial Management

Services:

Service	Primary Function	When to Use	Key Differentiator
AWS Billing Conductor	Customize billing logic	Complex enterprise billing scenarios	Allows custom billing logic creation
AWS Budgets	Set spending limits and alerts	Proactive cost control	Sets future limits, sends alerts
AWS Cost and Usage Report	Detailed cost data	Comprehensive cost analysis	Most granular cost data available
AWS Cost Explorer	Visualize and analyze costs	Understanding spending patterns	Visual reports and forecasting
AWS Marketplace	Third-party software catalog	Finding or selling AWS-compatible software	Not strictly for financial management

Common Mistakes to Avoid:

1. Don't confuse AWS Budgets with AWS Cost Explorer:
   - **Budgets is proactive (sets future limits)**
   - **Cost Explorer is for analysis and visualization of past and current costs**
2. AWS Cost and Usage Report vs. Cost Explorer:
   - Cost and Usage Report provides raw data
   - Cost Explorer provides visualizations and analysis tools
3. AWS Billing Conductor is not for individual accounts:
   - It's specifically for customizing billing in AWS Organizations
4. AWS Marketplace isn't just for buying:
   - It's also a platform for selling your own AWS-compatible solutions

Compute

1. AWS Batch
    - Key purpose: Run batch computing workloads on AWS
    - Use cases: • Data processing jobs • Machine learning model training
    - Exam tips: • Fully managed service • Automatically provisions compute resources based on volume and requirements of batch jobs
2. Amazon EC2 (Elastic Compute Cloud)
    - Key purpose: Resizable compute capacity in the cloud
    - Use cases: • Web servers • Application servers • Development and test environments
    - Exam tips: • Fundamental AWS service • Multiple instance types and pricing options (On-Demand, Reserved, Spot)
3. AWS Elastic Beanstalk
    - Key purpose: Easy-to-use service for deploying and scaling web applications
    - Use cases: • Deploying web applications without managing infrastructure
    - Exam tips: • Platform as a Service (PaaS) offering • Supports multiple languages and web servers
4. Amazon Lightsail
    - Key purpose: Simplified virtual private servers
    - Use cases: • Simple web applications • Websites for small businesses
    - Exam tips: • Easiest way to launch and manage a virtual private server on AWS • Fixed pricing plans
5. AWS Local Zones
    - Key purpose: Place compute, storage, and other select AWS services closer to end-users
    - Use cases: • Low-latency applications • Media and entertainment content creation
    - Exam tips: • Extension of an AWS Region • Located in large population centers
6. AWS Outposts
    - Key purpose: Run AWS infrastructure and services on-premises
    - Use cases: • Hybrid cloud deployments • Applications requiring low latency access to on-premises systems
    - Exam tips: • Brings AWS services to your data center • Fully managed by AWS
7. AWS Wavelength
    - Key purpose: Delivers ultra-low latency applications for 5G devices
    - Use cases: • Mobile edge computing applications • IoT applications requiring real-time processing

- Exam tips: • Embeds AWS compute and storage services within 5G networks • Designed for applications requiring ultra-low latency

Service	Primary Function	When to Use	Key Differentiator
AWS Batch	Batch computing workloads	Processing large volumes of data	Automatically scales compute resources
Amazon EC2	Resizable compute capacity	General purpose computing needs	Flexible, foundational compute service
AWS Elastic Beanstalk	Deploy and scale web applications	Simplified web app deployment	PaaS, handles infrastructure management
Amazon Lightsail	Simplified virtual private servers	Simple web applications, small businesses	Easy to use, fixed pricing plans
AWS Local Zones	Compute closer to end-users	Low-latency applications in specific areas	Extension of AWS Region in population centers
AWS Outposts	AWS services on-premises	Hybrid cloud, low-latency needs on-premises	Brings AWS to your data center
AWS Wavelength	Ultra-low latency for 5G	Mobile edge computing, IoT	Embedded in 5G networks

Key points:

1. EC2 is the foundational compute service, offering the most flexibility but requiring more management.
2. Elastic Beanstalk is for easier deployment of web applications, handling infrastructure management.
3. Lightsail is the simplest option for basic virtual private servers.
4. Batch is specifically for batch computing workloads.
5. Local Zones, Outposts, and Wavelength are about bringing AWS services closer to users or on-premises:
   - Local Zones: closer to population centers
   - Outposts: in your own data center
   - Wavelength: integrated with 5G networks

Containers

Service	Primary Function	When to Use	Key Differentiator
Amazon ECR	Container image registry	Storing and managing Docker images	Integrated with ECS and EKS
Amazon ECS	Container orchestration	Running Docker containers at scale	Native AWS service, deep integration
Amazon EKS	Managed Kubernetes	Running Kubernetes workloads	Kubernetes-specific, portability

Key points:

1. ECR is for storing and managing container images, not for running containers.
2. ECS is AWS's own container orchestration service, deeply integrated with other AWS services.
3. EKS is specifically for Kubernetes workloads, offering more portability but potentially more complexity.
4. All three services work together: you can store images in ECR and deploy them using either ECS or EKS.
5. **Fargate** can be used with both ECS and EKS for serverless container deployment.

Customer Engagement

Service	Primary Function	Target Audience	Key Differentiator
AWS Activate for Startups	Startup resources	Early-stage startups	Free credits and resources
AWS IQ	Expert marketplace	Customers needing project help	Pay-per-project AWS expertise
AWS Managed Services	Ongoing infrastructure management	Large enterprises	Full-scale AWS operations management
AWS Support	Tiered customer support	All AWS customers	Ranging from basic to enterprise-level support

Database

Amazon Aurora

- MySQL and PostgreSQL-compatible relational database
- Automatically grows storage as needed
- Designed for high availability and durability

Amazon DynamoDB

- Fully managed NoSQL database
- **Serverless** - no need to manage servers
- Scales automatically to handle massive workloads
- Used for applications needing consistent, single-digit millisecond latency

Amazon MemoryDB for Redis:

- Redis-compatible, durable, in-memory database service
- Ultra-fast performance with microsecond read latency
- Ideal for real-time applications requiring fast data access

Amazon Neptune

- Fully managed graph database service
- Optimized for storing billions of relationships
- Used for social networking, recommendation engines, fraud detection

Amazon RDS (Relational Database Service)

- **Managed** relational database service for multiple database engines
- Supports MySQL, PostgreSQL, MariaDB, Oracle, SQL Server
- Automates time-consuming administration tasks like hardware provisioning, patching
- Offers Multi-AZ deployment for high availability

Developer Tools

1. **AWS AppConfig** • Key purpose: Manage application configurations • Exam tips:
    - Allows dynamic updates to application configurations without redeploying
    - Useful for feature flags, application tuning, and A/B testing
2. **AWS CLI** (Command Line Interface) • Key purpose: Manage AWS services from the command line • Exam tips:
    - Unified tool to manage multiple AWS services
    - Alternative to using AWS Management Console

3. **AWS Cloud9** • Key purpose: Cloud-based IDE (Integrated Development Environment) • Exam tips:
    - Write, run, and debug code in your browser
    - Collaborate with others in real-time
4. **AWS CloudShell** • Key purpose: Browser-based shell environment • Exam tips:
    - Pre-authenticated AWS CLI access
    - No need to install or configure anything on your local machine
5. **AWS CodeArtifact** • Key purpose: Artifact repository service • Exam tips:
    - Store, publish, and share software packages
    - Integrates with common package managers like npm, pip, and Maven
6. **AWS CodeBuild** • Key purpose: Fully managed build service • Exam tips:
    - Compiles source code, runs tests, and produces software packages
    - Pay only for the compute time you use
7. **AWS CodeCommit** • Key purpose: Managed source control service • Exam tips:
    - Git-based repositories
    - Secure and scalable source control
8. **AWS CodeDeploy** • Key purpose: Automated deployment service • Exam tips:
    - Deploys to EC2, on-premises instances, Lambda functions, or ECS services
    - Minimizes downtime during deployments
9. **AWS CodePipeline** • Key purpose: Continuous delivery service • Exam tips:
    - Automates the build, test, and deploy phases of release process
    - Integrates with other AWS and third-party services

10. **AWS CodeStar** • Key purpose: Unified interface for software development • Exam tips:
    - Provides a dashboard for entire software development workflow
    - Quickly set up entire continuous delivery toolchain
11. **AWS X-Ray** • Key purpose: Analyze and **debug** distributed applications • Exam tips:
    - Provides view of request's path through your application
    - Helps identify performance bottlenecks and **troubleshoot** issues

Key Differentiators:

- CodeCommit (source control) → CodeBuild (build and test) → CodeDeploy (deployment) → CodePipeline (orchestrates the entire process)
- CodeStar is an overarching service that sets up a complete development toolchain
- Cloud9 is for writing code, while CloudShell is for running AWS CLI commands
- AppConfig is specifically for managing application configurations
- X-Ray is for application performance analysis and debugging

End User Computing

Service	**Primary Function**	**Use Case**	**Key Advantage**
AppStream 2.0	Application streaming	Access specific applications	No application redesign needed
WorkSpaces	Full desktop experience	Remote work, consistent desktops	Managed VDI solution
WorkSpaces Web	Web-based workspace	Secure access to internal sites/SaaS	Lightweight, browser-based access

Key points to remember for the exam:

1. AppStream 2.0 is for streaming specific applications, not full desktops.
2. WorkSpaces provides a full, persistent desktop experience.
3. WorkSpaces Web is specifically for browser-based access to web content and applications.
4. All these services help reduce the need for on-premises infrastructure management.
5. They all follow a pay-as-you-go model, aligning with AWS's cost-effective approach.

Frontend Web & Mobile

Service	Primary Function	Use Case	Key Advantage
AWS Amplify	Full-stack app development	Building web and mobile apps quickly	Integrated platform for frontend and backend
AWS AppSync	GraphQL API development	Real-time and offline-enabled apps	Simplified data synchronization
AWS Device Farm	App testing	Ensuring app quality across devices	Access to real devices in the cloud

Key points to remember for the exam:

1. AWS Amplify is a comprehensive platform for building full-stack applications quickly.
2. AWS AppSync focuses on GraphQL API development and real-time data synchronization.
3. AWS Device Farm is specifically for testing applications on real devices in the cloud.
4. Amplify can work with AppSync for GraphQL API integration in full-stack apps.
5. All these services aim to simplify and accelerate the development and delivery of high-quality web and mobile applications.

IoT

Service	Primary Function	Use Case	Key Advantage
AWS IoT Core	Cloud-based IoT device management	Large-scale IoT deployments	Secure, scalable device connectivity
AWS IoT Greengrass	Edge computing for IoT	Local processing on IoT devices	Offline operation and local execution

Key points to remember for the exam:

1. AWS IoT Core is for cloud-based management and connectivity of IoT devices at scale.
2. AWS IoT Greengrass focuses on extending AWS capabilities to edge devices for local processing.
3. IoT Core is about connecting devices to the cloud, while Greengrass is about bringing cloud capabilities to devices.
4. Both services enhance security for IoT deployments.
5. Greengrass allows for offline operation and local execution of Lambda functions.

ML

Service	**Primary Function**	**Key Use Case**
Comprehend	NLP	Sentiment analysis, entity recognition
Kendra	Intelligent search	Internal document search
Lex	Conversational AI	Chatbots, voice assistants
Polly	Text-to-Speech	Converting text to lifelike speech
Rekognition	Image/Video Analysis	Object detection, facial analysis
SageMaker	ML Platform	Building and deploying ML models
Textract	Document text extraction	Extracting data from forms, tables
Transcribe	Speech-to-Text	Converting audio to text
Translate	Language Translation	Translating text between languages

Management and Governance

1. **AWS Auto Scaling**: Automatically adjusts resources based on demand.
    - Key words: dynamic scaling, performance optimization

2. **AWS CloudFormation**: Infrastructure as Code (IaC) service.
    - Key words: templates, stacks, repeatable deployments
3. **AWS CloudTrail**: Logs API activity for auditing and compliance.
    - Key words: API calls, account activity, security analysis
4. **Amazon CloudWatch**: Monitoring and observability service.
    - Key words: metrics, alarms, logs, events
5. **AWS Compute Optimizer**: Analyzes resource usage and provides recommendations.
    - Key words: cost reduction, performance improvement, AI-powered
6. **AWS Config**: Assesses, audits, and evaluates resource configurations.
    - Key words: compliance, resource inventory, configuration history
7. **AWS Control Tower**: Sets up and governs multi-account AWS environments.
    - Key words: landing zone, guardrails, account factory
8. **AWS Health Dashboard**: Provides status of AWS services and account health.
    - Key words: personalized view, service health, planned changes
9. **AWS Launch Wizard**: Simplifies sizing, configuration, and deployment of third-party applications.
    - Key words: guided deployments, best practices, reduced time-to-deploy
10. **AWS License Manager**: Manages software licenses across AWS and on-premises.
    - Key words: license tracking, rule-based controls, reduce overages
11. **AWS Management Console**: Web interface to access and manage AWS services.
    - Key words: user-friendly, centralized management, web-based

12. **AWS Organizations**: Centrally manage and govern multiple AWS accounts.
    - Key words: consolidated billing, service control policies (SCPs), organizational units (OUs)
13. **AWS Resource Groups and Tag Editor**: Organize and manage AWS resources.
    - Key words: grouping, tagging, resource management
14. **AWS Service Catalog:** Create and manage catalogs of approved IT services.
    - Key words: self-service portal, standardized deployments, governance
15. **AWS Systems Manager**: Operational hub for AWS and on-premises resources.
    - Key words: patch management, automation, parameter store
16. **AWS Trusted Adviso**r: Provides real-time guidance on AWS best practices.
    - Key words: cost optimization, performance, security, fault tolerance
17. **AWS Well-Architected Tool:** Reviews workloads against AWS architectural best practices.
    - Key words: six pillars, workload reviews, improvement plans

Example scenario: "A company wants to ensure their AWS resource configurations comply with internal policies and industry regulations. Which service should they use?"

Approach:

1. Identify the problem: compliance and resource configuration management
2. Key words: resource configurations, comply, policies, regulations
3. Match with service descriptions: AWS Config fits this scenario perfectly

Migration and Transfer

1. **AWS Application Discovery Service:** Collects information about on-premises applications for planning migrations.
    - Key words: discovery, inventory, dependency mapping, on-premises
2. **AWS Application Migration Service (MGN):** Automates lift-and-shift migration of applications to AWS.
    - Key words: lift-and-shift, automated replication, minimal downtime
3. **AWS Database Migration Service (DMS):** Migrates databases to AWS with minimal downtime.
    - Key words: database migration, heterogeneous databases, continuous replication
4. **AWS Migration Hub**: Provides a single location to track migration progress across multiple AWS tools.
    - Key words: central dashboard, migration tracking, progress visualization
5. **AWS Schema Conversion Tool (SCT):** Converts database schemas from one engine to another.
    - Key words: schema conversion, database engine migration, code conversion
6. **AWS Snow Family:** Physical devices for secure, large-scale data transfer and edge computing.
    - Snowcone: Smallest device, 8 TB storage
    - Snowball: Mid-size device, 80 TB storage
    - Snowmobile: Exabyte-scale transfer using a shipping container
    - Key words: offline data transfer, edge computing, rugged environments
7. **AWS Transfer Family:** Managed file transfer services supporting SFTP, FTPS, and FTP protocols.

  - Key words: secure file transfer, SFTP, FTPS, FTP, existing workflows

tips for Migration and Transfer services:

1. Application Discovery Service is often used before Application Migration Service to plan migrations.
2. Migration Hub integrates with other AWS migration services to provide a centralized view.
3. Snow Family is ideal for situations with limited network bandwidth or very large amounts of data.
4. Transfer Family is useful when maintaining existing file transfer workflows while moving to AWS.

Example scenario: "A company wants to migrate their on-premises Oracle database to Amazon Aurora PostgreSQL with minimal downtime. Which two services should they use together?"

Approach:

1. Identify the problem: database migration from Oracle to PostgreSQL with minimal downtime
2. Key words: database migration, schema conversion, minimal downtime
3. Match with service descriptions:
   - AWS Database Migration Service (DMS) for minimal downtime migration
   - AWS Schema Conversion Tool (SCT) for converting Oracle schema to PostgreSQL
4. Eliminate others: Application Discovery Service (for app discovery), Snow Family (for large-scale data transfer)

Answer: AWS Database Migration Service (DMS) and AWS Schema Conversion Tool (SCT)

Networking & Content Delivery

1. **Amazon API Gateway**: Managed service for creating, publishing, and securing APIs.
   - Key words: RESTful APIs, WebSocket APIs, API management, serverless
2. **Amazon CloudFront**: Global content delivery network (CDN) service.

    - Key words: low-latency content delivery, edge locations, caching, DDoS protection
3. **AWS Direct Connect:** Dedicated private network connection from on-premises to AWS.
    - Key words: private connectivity, consistent network performance, reduced bandwidth costs
4. **AWS Global Accelerator**: Improves availability and performance using the AWS global network.
    - Key words: static IP addresses, fast regional failover, TCP/UDP applications
5. **Amazon Route 53**: Scalable domain name system (DNS) web service.
    - Key words: domain registration, DNS routing, health checking, latency-based routing
6. **Amazon VPC** (Virtual Private Cloud): Isolated section of the AWS Cloud for launching resources.
    - Key words: private network, subnets, network access control, security groups
7. **AWS VPN**: Encrypted connection between on-premises networks and AWS VPCs.
    - Key words: Site-to-Site VPN, Client VPN, encrypted connectivity, public internet

Example scenario: "A company wants to improve the global performance and availability of their web application while using their existing domain name. Which two services should they use together?"
Approach:

1. Identify the problem: global performance, availability, using existing domain
2. Key words: global performance, availability, domain name
3. Match with service descriptions:
    - CloudFront for global content delivery and improved performance
    - Route 53 for DNS management and routing using the existing domain
4. Eliminate others: Direct Connect (private connectivity), VPC (regional networking)

Answer: Amazon CloudFront and Amazon Route 53

Additional tips for Networking and Content Delivery services:

1. API Gateway is often used with Lambda for serverless architectures.
2. CloudFront works with WAF for additional security at the edge.
3. Direct Connect is ideal for consistent, high-bandwidth requirements.
4. Global Accelerator is useful for non-HTTP/S applications requiring static IP addresses.
5. VPC is the foundation for most AWS networking scenarios.
6. VPN is a cost-effective solution for secure connectivity over the public internet.

Key comparisons to remember:

- CloudFront vs Global Accelerator:
    - CloudFront: Content caching, best for static content and websites
    - Global Accelerator: IP address stability, best for dynamic content and non-HTTP applications
- Direct Connect vs VPN:
    - Direct Connect: Dedicated, private connection, higher bandwidth, more consistent
    - VPN: Encrypted connection over public internet, quicker to set up, more flexible
- Route 53 routing policies:
    - Simple, Weighted, Latency-based, Geolocation, Failover, Multivalue answer

Key services for on-premises to AWS connectivity:

1. AWS Direct Connect 2. AWS VPN (Site-to-Site VPN) 3. AWS VPN CloudHub 4. AWS Direct Connect + VPN

1. AWS Direct Connect:
    - Purpose: Dedicated private network connection
    - Key features: High bandwidth, consistent performance, reduced data transfer costs
    - Exam tricks to watch for:
        - Questions implying immediate setup (Direct Connect takes weeks to establish)
        - Scenarios requiring encryption (Direct Connect doesn't encrypt by default)
2. AWS Site-to-Site VPN:
    - Purpose: Encrypted connection over the public internet
    - Key features: Quick to set up, uses IPsec, works with existing VPN equipment
    - Exam tricks to watch for:
        - Questions suggesting it's a private connection (it uses the public internet)

        - Scenarios requiring very high, consistent bandwidth (VPN can be affected by internet congestion)

3. AWS VPN CloudHub:
    - Purpose: Connect multiple on-premises sites in a hub-and-spoke model
    - Key features: Uses Virtual Private Gateway, cost-effective for multiple sites
    - Exam tricks to watch for:
        - Questions that might confuse this with Direct Connect Gateway (different services)
4. AWS Direct Connect + VPN:
    - Purpose: Combines dedicated bandwidth of Direct Connect with VPN encryption
    - Key features: High security, consistent performance, encrypted traffic
    - Exam tricks to watch for:
        - Questions that imply this is a standard offering (it's a combined solution)

Security, Identity, and Compliance

Group services by function:

1. **Identity and Access**: IAM, IAM Identity Center, Cognito, Directory Service
2. **Encryption and Key Management**: KMS, CloudHSM, Certificate Manager
3. **Threat Detection and Monitoring:** GuardDuty, Detective, Inspector, Macie
4. **Network Security**: WAF, Shield, Network Firewall, Firewall Manager
5. **Compliance and Auditing**: Artifact, Audit Manager, Security Hub
6. **Resource Sharing:** Resource Access Manager (RAM)
7. **Secrets Management**: Secrets Manager

1. **Identity and Access Management:**

- **AWS IAM**: Manage access to AWS services and resources
    - Key point: Used for AWS users and resources
    - Trap: Don't confuse with on-premises identity management
- **AWS IAM Identity Center** (Single Sign-On): Centrally manage access to multiple AWS accounts and applications
    - Key point: SSO for AWS accounts and business applications
    - Trap: Don't confuse with Cognito (for customer-facing apps)
- **Amazon Cognito**: Add user sign-up, sign-in, and access control to web and mobile apps
    - Key point: For customer-facing applications
    - Trap: Not for AWS console or internal enterprise access
- **AWS Directory Service**: Managed Microsoft Active Directory
    - Key point: For Microsoft AD in the cloud
    - Trap: Not the same as IAM for AWS resource access

2. **Encryption and Key Management:**

- **AWS KMS**: Managed service to create and control encryption keys
    - Key point: For most AWS encryption needs
    - Trap: Don't confuse with CloudHSM for specific compliance requirements

• **AWS CloudHSM**: Hardware security modules for regulatory compliance

- Key point: Dedicated hardware, customer-managed
- Trap: Not the default choice for most encryption needs

• **AWS Certificate Manager**: Provision, manage, and deploy SSL/TLS certificates

- Key point: For securing websites and applications
- Trap: Not for internal PKI or non-AWS resources

3. **Threat Detection and Monitoring:**

• **Amazon GuardDuty**: Intelligent **threat detection** for AWS **accounts and workloads**

- Key point: Analyzes multiple data sources
- Trap: Don't confuse with Inspector (for vulnerability assessment)

• **Amazon Detective**: Analyze and **visualize** security data to investigate potential issues

- Key point: For **root cause** analysis
- Trap: Not a real-time threat detection service

• **Amazon Inspector: Automated** security assessment service

- Key point: For finding vulnerabilities and deviations from **best practices**
- Trap: Don't confuse with GuardDuty (for active threat detection)

• **Amazon Macie**: Discover and protect **sensitive data**

- Key point: Focuses on PII and sensitive data in S3
- Trap: Not a general-purpose security service

4. **Network Security**:

• **AWS WAF**: Protects web applications from common exploits

- Key point: Application layer (Layer 7) protection
- Trap: Don't confuse with Shield (for DDoS protection)

• **AWS Shield**: DDoS protection

- Key point: Network and transport layer protection
- Trap: Doesn't provide application-layer protection like WAF

• **AWS Network Firewall**: Network security across VPCs and accounts

- Key point: For customizable network-level protection
- Trap: Not the same as security groups or NACLs

• **AWS Firewall Manager**: Centrally manage firewall rules

- Key point: For managing WAF, Shield, and Network Firewall across accounts
- Trap: Doesn't replace the need for individual firewall services

5. **Compliance and Auditing:**

• **AWS Artifact**: Self-service portal for on-demand access to AWS compliance reports

- ○ Key point: For accessing AWS compliance documents
- ○ Trap: Not for generating your own compliance reports

• **AWS Audit Manager**: Continuously audit AWS usage for simplified risk and compliance assessment

- ○ Key point: For ongoing audits and assessments
- ○ Trap: Don't confuse with Artifact (for AWS compliance docs)

• **AWS Security Hub**: Comprehensive view of security alerts and posture

- ○ Key point: Central place to manage security across accounts
- ○ Trap: Doesn't replace individual security services

6. **Resource Sharing:**

• **AWS Resource Access Manager (RAM)**: Securely share AWS resources across accounts

- ○ Key point: For multi-account resource sharing
- ○ Trap: Not for identity federation or access management

7. **Secrets Managemen**t:

• **AWS Secrets Manager**: Rotate, manage, and retrieve secrets

- ○ Key point: For database credentials, API keys, etc.
- ○ Trap: Don't confuse with Parameter Store (part of Systems Manager)

Serverless (Fargate & Lambda)

AWS Lambda:
• Purpose: Run code without provisioning or managing servers

- Supports multiple programming languages
- Automated scaling
- Pay only for compute time consumed
- Integrates with many AWS services

Use cases:

- Short-lived processes (up to 15 minutes)
- **Event-driven** applications
- Backend for web, mobile, IoT applications

AWS Fargate:
• Purpose: Run containers without managing servers or clusters

- Works with ECS (Elastic Container Service) and EKS (Elastic Kubernetes Service)
- No need to provision EC2 instances
- Automated scaling and patching
- Pay for vCPU and memory resources allocated

• Use cases:

- Long-running processes
- Microservices architecture
- Application migration to containers

Tips:

- Lambda is for short-lived processes, Fargate for longer ones
- If the scenario mentions processes over 15 minutes, lean towards Fargate
- Lambda scales per request, Fargate at the task level
- for event-driven architectures involving AWS services, Lambda is often the go-to

Storage

1. **Amazon Elastic Block Store (Amazon EBS):**

• Purpose: Persistent block-level storage for EC2 instances

- Attached to a single EC2 instance at a time
- Different volume types (gp2, gp3, io1, io2, st1, sc1)
- Snapshots for backups

• tip: it's for single EC2 instance attachment and persists independently of EC2 instance lifecycle

2. **Amazon Elastic File System (Amazon EFS):**

• Purpose: Scalable, elastic file storage for EC2 instances

- Can be mounted on multiple EC2 instances simultaneously
- Grows and shrinks automatically
- Supports NFS protocol

• tip: Think of EFS when you need shared file storage across multiple EC2 instances

3. **Amazon FSx:**

• Purpose: Fully managed file systems for widely-used file systems

• Variants:

- FSx for Windows File Server
- FSx for Lustre (high-performance computing)
- FSx for NetApp ONTAP
- FSx for OpenZFS

• tip: Choose FSx when you need a specific file system type (e.g., Windows shares, high-performance computing)

4. **Amazon S3 (Simple Storage Service):**

• Purpose: Object storage for the internet

- Unlimited storage
- 99.999999999% durability
- Various storage classes (Standard, Intelligent-Tiering, One Zone-IA, Glacier, etc.)

• tip: Default choice for scalable, durable object storage. Remember it's not for file systems or block storage

5. **Amazon S3 Glacier**:

• Purpose: Low-cost archive storage for long-term, infrequently accessed data

  - Various retrieval options (minutes to hours)
  - Glacier Instant Retrieval, Flexible Retrieval, and Deep Archive

6. **AWS Storage Gateway**:

• Purpose: Hybrid cloud storage, connecting on-premises to cloud storage

• Types:

  - File Gateway
  - Volume Gateway
  - Tape Gateway

• tip: Think of Storage Gateway for hybrid scenarios, connecting on-premises applications with AWS storage

7. **AWS Backup**:

• Purpose: Centralized backup service across AWS services

  - Policy-based backup solution
  - Supports various AWS resources (EC2, EBS, RDS, DynamoDB, etc.)

• tip: Choose AWS Backup for centralized, managed backup across multiple AWS services

8. **AWS Elastic Disaster Recovery** (formerly CloudEndure Disaster Recovery):

• Purpose: Quickly and easily recover your physical, virtual, and cloud-based servers into AWS

  - Continuous replication
  - Automated disaster recovery drills

• tip: Think of this for disaster recovery scenarios, especially when migrating on-premises workloads to AWS

Key Differences:

1. EBS vs EFS:
   - Tip: EBS is for single EC2 instance, EFS for multiple instances sharing files
2. AWS Backup vs Elastic Disaster Recovery:
   - Tip: AWS Backup for regular backups, Elastic Disaster Recovery for full site/server recovery
3. FSx Variants:
   - Tip: Windows File Server for Windows workloads, Lustre for high-performance computing
4. Storage Gateway Types:
   - Tip: File Gateway for file interface, Volume Gateway for iSCSI volumes, Tape Gateway for virtual tapes

Example Scenario: "A company needs to store large amounts of data that will be accessed infrequently but requires immediate access when needed. Which storage solution is most appropriate?"

Approach:

1. Large amounts of data - suggests object storage
2. Infrequently accessed - points towards a lower-cost tier
3. Immediate access when needed - rules out Glacier (except Glacier Instant Retrieval)

Answer: Amazon S3 with S3 Standard-Infrequent Access (S3 Standard-IA) storage class

Review

Amazon Route 53: route end user to apps hosted on AWS (Connect user requests to infrastructure in AWS and outside of AWS) - DNS for domain names

Amazon Augmented AI (Amazon A2I): create human review workflows for ML

AWS Cost Explorer: create reports
AWS Budgets: set alerts

S3 Intelligent-Tiering storage class: a. S3 Standard (frequent access tier) b. S3 Standard-IA (infrequent access tier)
Scenario-based example: Imagine you're managing a photo-sharing application. New photos are initially stored in S3 Standard. If a photo isn't viewed for 30 days, it's automatically moved to S3 Standard-IA. If someone views the photo later, it's moved back to S3 Standard. This process optimizes storage costs based on actual usage patterns.

CloudTrail: history activity & API calls
CloudWatch: monitoring

AWS Direct Connect: private connection between on-premises and AWS

Exam Tip: When you see questions about connecting on-premises to AWS:

- If it mentions "dedicated" or "private" connection, think AWS Direct Connect
- If it's about general internet access to/from VPC, think Internet Gateway
- If it's about faster content delivery to users, think CloudFront
- If it mentions VPN or secure connection over the internet, think Virtual Private Gateway

AWS CloudFormation: Infrastructure as Code (IaC) service for provisioning AWS resources

"A company needs to deploy identical network configurations in three different AWS regions. Which service should they use?" Answer: AWS CloudFormation

AWS Outposts

Key Function:

- Extends AWS infrastructure and services to on-premises locations

Networking Relevance:

- Creates a hybrid environment, connecting on-premises networks with AWS cloud
- Requires network connectivity between Outposts and AWS Region

Exam Tip: When questions mention running AWS services on-premises or needing low-latency access to on-premises systems while using AWS services, think AWS Outposts.

Scenario Example: "A manufacturing company needs to run AWS services with single-digit millisecond latency to control their factory equipment. Which AWS service should they consider?" Answer: AWS Outposts

Memory Aid: "OUT-posts" - Brings AWS OUTside of its usual locations and into your premises

EC2 Data & Storage:
Types: Instance Store, Amazon EBS (Elastic Block Store), S3 (Not attachable to EC2)

Use cases:

- Temporary processing → Instance Store
- OS or database storage → EBS

Cost consideration:

- Instance Store → Included in EC2 cost
- EBS → Pay for provisioned capacity

Persistence:

- If data needs to persist → EBS or S3
- If data can be lost → Instance Store

Global Infra:

Availability Zone (AZ) vs Region vs Edge Location

Exam Tips:

1. When you see "isolated portion" or "fault isolation" in relation to AWS infrastructure, think Availability Zone.
2. If a question mentions "geographical area" or "multiple isolated locations," it's likely referring to a Region.
3. For content delivery and caching, especially with CloudFront, think Edge Location.

Memory Aid: "AZ for Isolation, Region for Geography, Edge for Delivery"

Potential Scenario-Based Questions:

1. "A company wants to deploy their application with high availability within a single geographical area. What's the minimum number of [BLANK] they should use?" Answer: Availability Zones (minimum of two)
2. "An e-commerce company wants to reduce latency for their global user base. Which AWS infrastructure component should they leverage?" Answer: Edge Locations (for CloudFront content delivery)
3. "To comply with data sovereignty laws, a company needs to ensure their data remains within a specific country. Which AWS concept should they focus on when deploying their infrastructure?" Answer: Region (as Regions are country or area-specific)

Remember:

- AZs are about isolation and fault tolerance within a Region
- Regions are about geographical separation and data sovereignty
- Edge Locations are about content delivery and reducing latency

For the exam, always consider the context of the question. If it's talking about high availability or fault tolerance within a close area, it's likely referring to Availability Zones. If it's about broader geographical concerns or completely separate infrastructures, it's probably about Regions. And if it's discussing content delivery speed or global reach, think Edge Locations.

EC2 Intances Pricings & Plans

Comparison Framework:

Pricing Option	Discount	Flexibility	Commitment	Best For
On-Demand	None	Highest	None	Unpredictable workloads
Standard RI	Up to 72%	Low	1 or 3 years	Steady, predictable use
Convertible RI	Up to 54%	Medium	1 or 3 years	Long-term, changing needs
Savings Plans	Up to 72%	High	1 or 3 years	Consistent usage, need flexibility
Spot	Up to 90%	Highest (but interruptible)	None	Fault-tolerant, flexible timing
Dedicated Hosts	Variable	Low	On-Demand or Reserved	Compliance, licensing needs

Exam Tips:

1. Always consider the workload characteristics mentioned in the question:
   - Predictable, steady → Reserved Instances
   - Flexible, long-term → Savings Plans or Convertible RIs
   - Interruptible, cost-sensitive → Spot Instances
   - Short-term, cannot be interrupted → On-Demand
   - Compliance/licensing issues → Dedicated Hosts
2. Pay attention to keywords:
   - **"Hourly commitment" → Savings Plans**
   - "Specific instance type" → Standard Reserved Instances
   - "Flexible instance needs" → Convertible RIs or Savings Plans
   - "Highest discount" → Spot Instances (with caveat of interruptions)
   - Reserved Instances do not require an hourly spend commitment over the duration of the contract term.

Potential Exam Scenarios:

1. "A company runs a critical application that cannot be interrupted and has unpredictable usage patterns. Which pricing model is best?" Answer: On-Demand Instances
2. "An organization wants to run a batch processing job that can be interrupted. They need the most cost-effective solution. What should they choose?" Answer: Spot Instances
3. "A business has a stable workload but wants flexibility to change instance types as needed over the next year. Which option provides this flexibility with cost savings?" Answer: EC2 Instance Savings Plans or Convertible Reserved Instances
4. "A company needs to use its own software licenses and have control over instance placement. What EC2 option should they consider?" Answer: Dedicated Hosts

Final Memory Hook: "On-Demand for flexibility, Reserved for stability, Savings Plans for adaptability, Spot for volatility, and Dedicated for exclusivity"

A company has deployed applications on Amazon EC2 instances. The company needs to assess application vulnerabilities and must identify infrastructure deployments that do not meet best practices.
Which AWS service can the company use to meet these requirements?
Amazon Inspector!

AWS Fargate: A technology that you can use with Amazon ECS to run containers without having to manage servers or clusters of Amazon EC2 instances.

AWS Athena: An interactive query service that makes it simple to analyze data directly in Amazon S3 using standard SQL.

Which AWS services or tools can identify rightsizing opportunities for Amazon EC2 instances?
AWS Cost Explorer provides cost management tools that allow you to analyze your AWS spending, including EC2 instance usage. While it doesn't directly provide rightsizing recommendations, it offers insights into your usage patterns, which can inform rightsizing decisions.

AWS Compute Optimizer analyzes your Amazon EC2 usage patterns and provides recommendations for rightsizing your EC2 instances, helping you optimize performance and reduce costs. It considers factors such as CPU utilization, memory utilization, and network throughput to make recommendations tailored to your workload.

Benefits of **Trusted Advisor**:
• Cost optimization - Trusted Advisor can help you save cost with actionable recommendations by analyzing usage, configuration and spend.
• Performance - Trusted Advisor can help improve the performance of your services with actionable recommendations by analyzing usage and configuration.
• Security - Trusted Advisor can help improve the security of your AWS environment by suggesting foundational security best practices curated by security experts.
• Fault tolerance - Trusted Advisor can help improve the reliability of your services.
• Service quotas - Service quotas are the maximum number of resources that you can create in an AWS account.

AWS Service Catalog is the service that enables a company to manage deployed IT services and **govern** its infrastructure as code (IaC) templates. AWS Service Catalog allows organizations to create and manage catalogs of IT services that are approved for use on AWS. It helps centralize and manage the provisioning of resources and applications based on predefined templates.

AWS Glue is a fully managed extract, transform, and load (**ETL**) service that makes it easy to discover, prepare, and load data for analysis. It automates the time-consuming tasks of data discovery, transformation, and job scheduling, allowing users to focus on analyzing the data.

B. Amazon Elastic File System (Amazon EFS): Fully managed file system for shared access to file-based data. Not primarily used for data discovery, transformation, or visualization.

C. Amazon Redshift: Fully managed data warehouse service for running complex queries on large datasets. Not used for data discovery, transformation, or visualization.

D. Amazon **QuickSight**: Amazon QuickSight is a fully managed business intelligence (BI) service that enables users to create and visualize interactive dashboards and reports. It connects to various data sources, making it suitable for visualizing data prepared by services like AWS Glue.

Amazon Polly is a machine learning service that converts text to speech. This service provides the ability to read text out loud.

Developer tools:

● AWS AppConfig ● AWS CLI ● AWS Cloud9 ● AWS CloudShell ● AWS CodeArtifact ● AWS CodeBuild ● AWS CodeCommit ● AWS CodeDeploy ● AWS CodePipeline ● AWS CodeStar ● AWS X-Ray

CodeArtifact is a managed artifact repository service that stores and shares software that is ready for deployment. CodeArtifact is not a source code management service.

CodeBuild is a service that helps users to automatically compile source code, run unit tests, and produce software packages that are ready for deployment. CodeBuild is not a code management service.

CodePipeline is a service that manages the movement of code between the individual services. CodePipeline is not a source code storage service.

CodeCommit is a source code version control service. CodeCommit helps users store and manage developers' source code in AWS.

What are the advantages of deploying an application with Amazon EC2 instances in multiple Availability Zones?
prevents a single point of failure
f you host all your instances in a single location that is affected by a failure, none of your instances would be available. Availability Zones are designed for physical redundancy and to provide resilience with uninterrupted performance.
The best option to serve users with low latency across Regions is to launch another EC2 instance in the second Region that is closer to the user's location.

A user needs to automatically discover, classify, and protect sensitive data stored in Amazon S3.
Which AWS service can meet these requirements?
Amazon Macie is an automated security assessment service that helps improve the security and compliance of applications deployed on AWS.

Also S3, Versioning in Amazon S3 is a means of keeping multiple variants of an object in the same bucket. You can use the S3 Versioning feature to preserve, retrieve, and restore every version of every object stored in your buckets.
Versioning-enabled buckets can help you recover objects from accidental deletion or overwrite. For example, if you delete an object, Amazon S3 inserts a delete marker instead of removing the object permanently.

A company requires an encrypted connection between the company's on-premises servers and AWS. The connection must use the company's existing internet connection.

Which solution will meet these requirements?

1. Incorrect. Direct Connect links your internal network to a Direct Connect location over a network connection. One end of the connection attaches to your on-premises router. The other end connects to a Direct Connect router. With this connection, you can bypass the ISPs in your network path. However, the company must use an existing internet connection in this scenario.

2. Incorrect. Amazon Connect is an omnichannel cloud contact center. Amazon Connect helps you provide customer service at a low cost. Amazon Connect uses an omnichannel design to provide a seamless experience across voice and chat for your customers and agents. Amazon Connect does not provide a network connection.

3. Incorrect. CloudFront is a web service that speeds up the distribution of your static and dynamic web content to your users. CloudFront delivers your content through a worldwide network of data centers known as edge locations. When a user requests content that you serve through CloudFront, the request is routed to the edge location that provides the lowest latency. However, CloudFront does not provide a network connection.

4. Correct. AWS Site-to-Site VPN creates an encrypted network path between your on-premises network and your AWS Cloud network. This connection between your on-premises network and your AWS Cloud network uses the internet.

Which AWS service identifies security groups that allow unrestricted access to a user's AWS resources?

Trusted Advisor checks security groups for rules that allow unrestricted access to a resource. Unrestricted access increases opportunities for malicious activity, such as hacking, denial-of-service attacks, or loss of data.
CloudWatch is a monitoring service that collects and tracks metrics for AWS resources. It does not identify security groups that allow unrestricted access.

A company is hosting a static website from a single Amazon S3 bucket.
Which AWS service will achieve lower latency and high transfer speeds?

1. Incorrect. Elastic Beanstalk is a service to deploy and scale web applications and services developed with common programming languages on automatically deployed infrastructure with capacity management, load balancing, auto scaling, and monitoring. Elastic

Beanstalk makes it easier to provision and support an application. Elastic Beanstalk does not reduce website latency.

2. Correct. CloudFront is a web service that speeds up the distribution of your static and dynamic web content, such as .html, .css, .js, and image files, to your users. Content is cached in edge locations. Content that is repeatedly accessed can be served from the edge locations instead of the source S3 bucket.

Which AWS service allows customers to purchase unused Amazon EC2 capacity at an often discounted rate?
With Spot Instances, you can access unused EC2 capacity. Spot Instances can be discounted.

With **Elastic Beanstalk**, you can quickly deploy and manage applications in the AWS Cloud without having to learn about the infrastructure that runs those applications. Elastic Beanstalk reduces management complexity without restricting choice or control. You simply upload your application, and Elastic Beanstalk automatically handles the details of capacity provisioning, load balancing, scaling, and application health monitoring.

A. AWS Support: Provides technical support plans, but not focused on hands-on migration assistance.

B. AWS Professional Services: Global team of experts for hands-on assistance with planning, executing, and optimizing AWS migrations.

C. AWS Launch Wizard: Simplifies application deployment, but not specifically designed for third-party application migrations.

D. AWS Managed Services (AMS): Fully managed service for ongoing operational support, not designed for the initial migration phase.

Security Mechanisim: Network ACL vs Security groups vs bucket policies - IAM
Budgets vs Cost explorer vs AWS Cost and Usage Reports vs Cost allocation tags
CloudTrail vs CloudWatch
CAF & WAF
AWS and on-premise concepts and servies