

MODULE-1:

Cloud Concepts Overview

Cloud computing is the on-demand delivery of computing power, storage, databases, and applications over the internet with a pay-as-you-go model. It allows users to treat infrastructure as software, enabling flexibility, scalability, and cost efficiency. Key cloud service models include:

1. **IaaS (Infrastructure as a Service):** Provides access to basic IT resources like virtual machines, storage, and networking.
2. **PaaS (Platform as a Service):** Allows users to focus on application deployment without managing the underlying infrastructure.
3. **SaaS (Software as a Service):** Delivers fully managed applications to end users.

Cloud deployment models include:

1. **Cloud:** Fully deployed in the cloud.
2. **Hybrid:** Combines on-premises and cloud resources.
3. **On-premises (private cloud):** Dedicated infrastructure managed by the organization.

Cloud computing is flexible, allowing users to provision, scale, and pay for resources based on demand, avoiding the limitations of traditional hardware-based solutions.

Here are the six key **advantages of cloud computing:**

1. **Trade Capital Expense for Variable Expense:** Instead of investing in expensive data centers and hardware, you only pay for the resources you use, saving money and increasing flexibility.
2. **Benefit from Massive Economies of Scale:** Cloud providers aggregate usage from many customers, allowing them to lower costs and pass on savings to users.
3. **Stop Guessing Capacity:** Scale up or down as needed without worrying about over- or underestimating resource needs.
4. **Increase Speed and Agility:** New IT resources are available in minutes, accelerating development and innovation.
5. **Stop Spending Money on Running Data Centers:** Focus on your core business rather than managing physical infrastructure.
6. **Go Global in Minutes:** Deploy applications worldwide easily, offering low-latency services to customers anywhere.

AWS Overview:

- **AWS (Amazon Web Services)** is a cloud platform offering on-demand access to computing, storage, networking, and databases.
- **Flexibility:** Resources can be scaled up or down as needed.
- **Pay-as-you-go:** You only pay for the services you use.

- **Service Categories:** Includes compute, storage, databases, networking, machine learning, and more.
- **Example Services:**
 - **Amazon EC2:** Full control over virtual servers.
 - **AWS Lambda:** Run code without managing servers.
 - **Amazon S3:** Object storage for any type of data.
- **Access Methods:**
 - **AWS Management Console** (GUI)
 - **AWS CLI** (Command Line)
 - **SDKs** (Programming Integration).

AWS Cloud Adoption Framework (AWS CAF) Overview (In Short):

- **Purpose:** AWS CAF provides guidance and best practices to help organizations adopt cloud computing effectively by aligning people, processes, and technology.
- **Six Core Perspectives:**
 - **Business:** Focuses on aligning IT and business strategies (stakeholders: business/finance managers).
 - **People:** Emphasizes training, staffing, and organizational change to build agility (stakeholders: HR and staffing).
 - **Governance:** Ensures IT strategy aligns with business goals and minimizes risks (stakeholders: CIOs, program managers).
 - **Platform:** Focuses on architecture and provisioning for cloud and on-premises workloads (stakeholders: CTOs, IT managers).
 - **Security:** Ensures security objectives, such as control and auditability, are met (stakeholders: CISOs, security analysts).
 - **Operations:** Defines operational processes and procedures for business support (stakeholders: operations/support managers).

Each perspective has a set of capabilities to identify areas requiring attention, helping streamline cloud adoption strategies.

| | |
|--|---|
| <ul style="list-style-type: none"> • Cloud computing is the on-demand delivery of IT resources via the internet with pay-as-you-go pricing. • Cloud computing enables you to think of (and use) your infrastructure as software. • There are three cloud service models: IaaS, PaaS, and SaaS. • There are three cloud deployment models: cloud, hybrid, and on-premises or private cloud. • Almost anything you can implement with traditional IT can also be implemented as an AWS cloud computing service. | <ul style="list-style-type: none"> • Trade capital expense for variable expense • Benefit from massive economies of scale • Stop guessing capacity • Increase speed and agility • Stop spending money on running and maintaining data centers • Go global in minutes |
| <ul style="list-style-type: none"> • AWS is a secure cloud platform that offers a broad set of global cloud-based products called services that are designed to work together. • There are many categories of AWS services, and each category has many services to choose from. • Choose a service based on your business goals and technology requirements. • There are three ways to interact with AWS services. | <ul style="list-style-type: none"> • Cloud adoption is not instantaneous for most organizations and requires a thoughtful, deliberate strategy and alignment across the whole organization. • The AWS CAF was created to help organizations develop efficient and effective plans for their cloud adoption journey. • The AWS CAF organizes guidance into six areas of focus, called perspectives. • Perspectives consist of sets of business or technology capabilities that are the responsibility of key stakeholders. |

Why is AWS more economical than traditional data centers for applications with varying compute workloads?

| Choice | Response |
|--------|--|
| A | Amazon Elastic Compute Cloud (Amazon EC2) costs are billed on a monthly basis. |
| B | Customers retain full administrative access to their Amazon EC2 instances. |
| C | Amazon EC2 instances can be launched on-demand when needed. |
| D | Customers can permanently run enough instances to handle peak workloads. |

The correct answer is C.

MODULE -2

FUNDAMENTALS OF PRICING

- **Pay for What You Use:** AWS operates on a pay-as-you-go model, allowing you to only pay for services consumed, eliminating large upfront costs and infrastructure investments.
- **Reserved Instances (RIs):** Investing in RIs can save you up to 75% compared to on-demand pricing. There are three options:
 - **All Upfront Reserved Instance (AURI):** Largest discount.
 - **Partial Upfront Reserved Instance (PURI):** Lower discount with partial upfront payment.
 - **No Upfront Payments Reserved Instance (NURI):** Smallest discount, no upfront costs.
- **Volume-Based Discounts:** AWS offers tiered pricing for services like Amazon S3, where costs decrease as usage increases.
- **Cost Reductions Over Time:** AWS consistently lowers prices and improves efficiencies, with significant price reductions since 2006.
- **Custom Pricing:** Available for high-volume projects with unique needs.
- **AWS Free Tier:** New customers can access a free usage tier for one year, including services like Amazon EC2 and Amazon S3.
- **No Additional Charge Services:** Certain AWS services (like VPC, IAM, Elastic Beanstalk) are available at no charge, but using them may incur costs for other associated services.

TOTAL COST OF OWNERSHIP:

Key Aspects of Total Cost of Ownership (TCO) and Cloud Migration

Definition of TCO:

1. TCO includes all costs associated with acquiring, operating, and maintaining a system throughout its lifecycle (e.g., acquisition, operational, maintenance).

Cost-Benefit Analysis:

1. Evaluate financial implications of cloud migration vs. on-premises infrastructure, considering both tangible and intangible benefits.

Operational Efficiency:

1. Cloud automation reduces routine tasks, allowing IT teams to focus on strategic initiatives, improving deployment speed.

Scalability and Flexibility:

1. Cloud resources can scale up or down based on demand, optimizing costs and performance without significant capital investments.

Enhanced Security and Compliance:

1. Cloud providers offer robust security measures and compliance frameworks that can be more efficient than in-house solutions.

Disaster Recovery:

1. Built-in disaster recovery capabilities ensure business continuity at a lower cost compared to traditional setups.

Pay-as-You-Go Model:

1. Organizations only pay for resources used, reducing over-provisioning and ensuring cost-effectiveness.

Long-term Strategy and ROI:

1. Migration should align with strategic goals, showing positive ROI not just in cost savings but also in revenue potential and innovation.

Conclusion

A thorough evaluation of TCO is essential for successful cloud migration, leading to improved efficiency, agility, and competitiveness in a digital landscape.

AWS ORGANIZATION :

Introduction to AWS Organizations

Overview:

- AWS Organizations is a free account management service that consolidates multiple AWS accounts for centralized management, including consolidated billing and account management.

Key Benefits:

- **Centralized Management:** Manage access policies across multiple AWS accounts.
- **Controlled Access:** Define access to AWS services.
- **Automated Account Management:** Streamline the creation and management of AWS accounts.
- **Consolidated Billing:** Simplify billing across all accounts and leverage volume discounts.

Terminology

- **Organizational Units (OUs):** Containers for accounts within a root, allowing for hierarchical structure and policy application.
- **Service Control Policies (SCPs):** Policies that control access to AWS services across accounts within OUs.

Key Features

- **Policy-based Account Management:** Apply policies at the OU level or directly to accounts.
- **APIs for Automation:** Automate account management tasks.
- **Consolidated Billing:** One payment method for all accounts, providing visibility into charges and cost management.

Security

- **IAM Policies vs. SCPs:**
 - IAM policies control access at the user level, while SCPs manage access at the account or OU level.
 - SCPs impact all users, including the root user of the account.

Setup Process

1. Create your organization with the primary account.
2. Create organizational units and place member accounts within them.
3. Create SCPs to apply restrictions.
4. Test the policies to verify access controls.

Limits of AWS Organizations

- Various limits exist on naming conventions, account numbers, organizational units, policies, and more.

Accessing AWS Organizations

- Managed through:
 - **AWS Management Console:** Browser-based interface.
 - **AWS CLI:** Command line for faster task execution.
 - **SDKs:** Libraries for various programming languages.
 - **HTTPS API:** Programmatic access to AWS Organizations.

AWS Billing and Cost Management :

AWS Billing Dashboard

Overview:

- The AWS Billing Dashboard provides insights into monthly AWS expenditures, helping users track and analyze their spending.

Key Features:

1. Spend Summary:

- Displays last month's spending, current month-to-date costs, and forecasts for the upcoming month.

2. Month-to-Date Spend by Service:

- Highlights the top services contributing to overall expenditure.

Cost Management Tools

From the billing dashboard, users can access several cost management tools:

AWS Budgets:

- Allows users to set and track budgets.
- Sends notifications when costs exceed set budgets.
- Budgets can be customized for monthly, quarterly, or yearly tracking.

AWS Cost Explorer:

- Visualizes AWS cost data over time.
- Offers reports on costs and usage for the past 13 months.
- Provides forecasts for the upcoming three months and identifies spending patterns.

AWS Cost and Usage Report:

- A comprehensive report detailing AWS costs and usage.
- Offers hourly or daily usage line items for each service.
- Can publish reports to an S3 bucket, updated daily.

Monthly Bills

- The **AWS Bills** page provides a detailed breakdown of costs incurred for each AWS service, including regional and linked account details.

AWS Support Plans

AWS offers four support plans to cater to different customer needs:

Basic Support:

1. Access to 24/7 customer service, documentation, whitepapers, and support forums.
2. Core Trusted Advisor checks.
3. Personal Health Dashboard.

Developer Support:

1. Resources for customers in early development stages.
2. Guidance and technical support for testing and deploying applications.
3. Suitable for non-production workloads.

Business Support:

1. Designed for customers running production workloads.
2. Offers resources for applications in production environments.
3. Provides additional technical support and faster response times.

Enterprise Support:

1. For business and mission-critical workloads.
2. Focuses on proactive management and operational efficiency.
3. Includes access to a Technical Account Manager (TAM) for tailored support.

Case Severity and Response Times

AWS support cases are categorized into five severity levels, influencing response times:

1. **Critical:** Business at risk; critical application functions unavailable.
2. **Urgent:** Significant business impact; important functions unavailable.
3. **High:** Important functions impaired or degraded.
4. **Normal:** Non-critical functions behaving abnormally; time-sensitive questions.
5. **Low:** General development questions or feature requests.

Note: The Basic Support Plan does not include case support.

Tools for Cost Management

- **AWS Pricing Calculator:** Helps estimate costs and model solutions.
- **AWS Billing Dashboard:** Provides tools for understanding and managing AWS costs, including:
 - **AWS Bills**
 - **AWS Cost Explorer**
 - **AWS Budgets**
 - **AWS Cost and Usage Reports**

Summary of the Module

- Reviewed AWS pricing fundamentals and Total Cost of Ownership (TCO) concepts.
- Explored various AWS tools to manage and optimize costs effectively.
- Encouraged understanding AWS service usage and associated costs for better planning.

- There is no charge (with some exceptions) for:
 - Inbound data transfer.
 - Data transfer between services within the same AWS Region.
- Pay for what you use.
- Start and stop anytime.
- No long-term contracts are required.
- Some services are free, but the other AWS services that they provision might not be free.

Which AWS service provides infrastructure security optimization recommendations?

| Choice | Response |
|----------|--|
| A | AWS Price List Application Programming Interface (API) |
| B | Reserved Instances |
| C | AWS Trusted Advisor |
| D | Amazon Elastic Compute Cloud (Amazon EC2) Spot Fleet |

The correct answer is C.

MODULE-3

AWS Global Infrastructure Overview:

AWS Global Infrastructure Overview

Regions and Availability Zones (AZs):

1. **22 Regions** worldwide, each with multiple **Availability Zones** for fault tolerance.
2. Data replication across Regions is the user's responsibility.

Factors for Selecting Regions:

1. **Data Governance:** Compliance with local laws.
2. **Proximity:** Reduces latency for users.
3. **Service Availability:** Not all services are available in every Region.
4. **Cost Differences:** Running costs can vary by Region.

Availability Zones:

1. Physically isolated but interconnected for high availability.
2. Recommended to replicate resources across AZs for resilience.

Data Centers:

1. Secure facilities designed for redundancy in power and networking.
2. AWS ensures service availability through constant monitoring.

Content Delivery:

1. **Amazon CloudFront** and **Route 53** reduce latency by routing requests to the nearest edge location.

Infrastructure Benefits:

1. **Elasticity:** Resources adjust to demand.
2. **Fault Tolerance:** Redundant components ensure continued operation.
3. **High Availability:** Minimal downtime with low human intervention.

Overview of AWS Services

1. AWS Database Services

- **Amazon RDS:** Simplifies setup and management of relational databases, automating tasks like backups and patching.

- **Amazon Aurora:** MySQL and PostgreSQL-compatible database, faster than standard versions (up to 5x MySQL and 3x PostgreSQL).
- **Amazon Redshift:** Data warehousing service for running analytics on large datasets, integrates with Amazon S3.
- **Amazon DynamoDB:** Key-value and document database offering single-digit millisecond performance, with built-in security and caching.

2. AWS Networking and Content Delivery Services

- **Amazon VPC:** Allows creation of isolated sections of the AWS cloud.
- **Elastic Load Balancing:** Distributes incoming traffic across various targets (e.g., EC2 instances).
- **Amazon CloudFront:** Fast CDN for delivering data globally with low latency.
- **AWS Transit Gateway:** Connects VPCs and on-premises networks.
- **Amazon Route 53:** DNS service for routing users to applications.
- **AWS Direct Connect:** Establishes dedicated private network connections to AWS.
- **AWS VPN:** Creates secure tunnels to the AWS global network.

3. AWS Security, Identity, and Compliance Services

- **AWS IAM:** Manages access to AWS services securely.
- **AWS Organizations:** Restricts services and actions across accounts.
- **Amazon Cognito:** Manages user sign-up and sign-in for apps.
- **AWS Artifact:** Access to security and compliance reports.
- **AWS KMS:** Key management service for encryption.
- **AWS Shield:** DDoS protection for applications on AWS.

4. AWS Cost Management Services

- **AWS Cost and Usage Report:** Comprehensive AWS cost and usage data.
- **AWS Budgets:** Custom budgets with alerts for spending.
- **AWS Cost Explorer:** Visualizes and manages AWS costs and usage over time.

5. AWS Management and Governance Services

- **AWS Management Console:** Web-based UI for AWS account access.
- **AWS Config:** Tracks resource inventory and changes.
- **Amazon CloudWatch:** Monitors resources and applications.
- **AWS Auto Scaling:** Scales resources based on demand.
- **AWS Command Line Interface:** Unified tool for managing AWS services.
- **AWS Trusted Advisor:** Optimizes performance and security.
- **AWS Well-Architected Tool:** Reviews and improves workloads.
- **AWS CloudTrail:** Tracks user activity and API usage.

- The **AWS Global Infrastructure** consists of **Regions** and **Availability Zones**.
- Your choice of a **Region** is typically based on **compliance requirements** or to **reduce latency**.
- Each **Availability Zone** is physically separate from other Availability Zones and has redundant power, networking, and connectivity.
- **Edge locations**, and **Regional edge caches** improve performance by **caching** content closer to users.

Which component of AWS global infrastructure does Amazon CloudFront use to ensure low-latency delivery?

| Choice | Response |
|----------|---|
| A | AWS Regions |
| B | AWS edge locations |
| C | AWS Availability Zones |
| D | Amazon Virtual Private Cloud (Amazon VPC) |

The correct answer is B.

MODULE-4

AWS shared responsibility model:

- **Customer Responsibilities:**

- • Manage operating system and applications on Amazon EC2, including patching and maintenance.
- Configure security groups, OS firewalls, and network settings.
- Handle account and access management, including passwords, IAM roles, and permissions.
- Manage client-side encryption, data integrity, and network traffic encryption.

- **Service Models:**

- • **IaaS (Infrastructure as a Service):**

- Customer controls networking, storage, and access configurations.
- Example: Amazon EC2, Amazon EBS.

- **PaaS (Platform as a Service):**

- AWS manages infrastructure, OS patching, and firewall configurations.
- Customer focuses on application management.
- Example: AWS Lambda, Amazon RDS.

- **SaaS (Software as a Service):**

- AWS provides fully managed software with customer access.
- No infrastructure management needed by the customer.
- Example: AWS Trusted Advisor, AWS Shield, Amazon Chime.

- **Scenario-Based Responsibilities:**

Scenario 1:

- EC2 OS patches, security groups, and application configurations are the customer's responsibility.
- Physical security, virtualization infrastructure, and RDS database patches are AWS's responsibility.

Scenario 2:

- Configuring subnets, VPC, SSH key security, and enabling MFA for users is the customer's responsibility.
- AWS manages network isolation, low-latency connections, and protection against regional network outages.

- AWS and the customer share security responsibilities:
 - AWS is responsible for security **of** the cloud
 - Customer is responsible for security **in** the cloud
- **AWS is responsible for protecting the infrastructure**—including hardware, software, networking, and facilities—that run AWS Cloud services
- For services that are categorized as infrastructure as a service (IaaS), the **customer is responsible for performing necessary security configuration and management tasks**
 - For example, guest OS updates and security patches, firewall, security group configurations

○

AWS Identity and Access Management (IAM):

- **IAM Overview:**

- IAM enables centralized access management for AWS resources.
- It's free and helps manage who can access which resources and what actions they can perform.

- **IAM Components:**

- **IAM Users:** Represents a person or application needing access to AWS services.
- **IAM Groups:** Groups of users with the same permissions, simplifying management.
- **IAM Policies:** JSON documents that define permissions. They dictate what actions are allowed or denied for users, groups, or roles.
- **IAM Roles:** Temporarily grant permissions for specific AWS resources.

- **Authentication Types:**

- **Programmatic Access:** Requires an Access Key ID and Secret Access Key, for CLI or SDK access.
- **Management Console Access:** Requires username, password, and optionally, Multi-Factor Authentication (MFA) for added security.

- **Multi-Factor Authentication (MFA):**

- Enhances security by requiring a unique authentication code in addition to username and password. Compatible with virtual, U2F, or hardware MFA devices.

- **Authorization:**

- By default, IAM entities have no permissions. Permissions are explicitly granted through policies, following the principle of least privilege.
- **Policy Types:**
 - **Identity-based:** Attached to IAM entities like users, groups, or roles.

- **Resource-based:** Attached directly to resources, like S3 buckets, controlling access at the resource level.

- **IAM policies** are constructed with JavaScript Object Notation (JSON) and define permissions.
 - IAM policies can be attached to any **IAM entity**.
 - Entities are IAM users, IAM groups, and IAM roles.
- An **IAM user** provides a way for a person, application, or service to authenticate to AWS.
- An **IAM group** is a simple way to attach the same policies to multiple users.
- An **IAM role** can have permissions policies attached to it and can be used to delegate temporary access to users or applications.

Securing a new AWS account :

AWS Account Root User vs. IAM Access

Root User:

1. **Complete Access:** Has full control over all AWS resources.
2. **Login:** Accessed with the email and password used to create the account.
3. **Critical Actions:**
 1. Update root password.
 2. Change AWS Support plan.
 3. Restore IAM user permissions.

Best Practices:

1. **Limit Root User Use:** Only use for critical tasks.
2. **Create IAM Users:** Follow the principle of least privilege.
3. **Unique Credentials:** Assign different permissions to each IAM user.

Securing Your AWS Account:

1. **Stop Using Root User:** Create an IAM user, remove root access keys.
2. **Enable MFA:** Require MFA for root and IAM user logins.
3. **Use AWS CloudTrail:** Monitor and log API activity.
4. **Enable Billing Reports:** Track usage and costs with AWS Cost and Usage Report.

Best practices to secure an AWS account:

- Secure logins with multi-factor authentication (MFA).
- Delete account root user **access keys**.
- Create individual **IAM users** and grant permissions according to the principle of least privilege.
- Use groups to assign permissions to IAM users.
- Configure a **strong password policy**.
- Delegate using **roles** instead of sharing credentials.
- Monitor account activity by using AWS CloudTrail.

Securing accounts :

AWS Organizations

- **Manage multiple accounts** centrally.
- **Security**: Use **OUs** for access policies and **SCPs** to set maximum permissions.

AWS Key Management Service (KMS)

- **Create/manage encryption keys**.
- **Integrates with CloudTrail** for logging and uses **validated HSMS** for protection.

Amazon Cognito

- **User authentication** for apps.
- Supports **social/enterprise logins** and complies with standards like **HIPAA** and **PCI DSS**.

AWS Shield

- **DDoS protection service**.
- **Standard**: Free for all AWS users.
- **Advanced**: Paid service for enhanced DDoS defense.

Securing data on AWS:

data encryption and securing Amazon S3:

Encryption of Data at Rest

- **Purpose**: Protects data by encoding it with a secret key.
- **Management**: AWS KMS manages encryption keys.
- **Data at Rest**: Refers to stored data (on disk/tape).
- **Services**: Supported services include Amazon S3, EBS, EFS, and RDS.
- **Encryption Standard**: Uses AES-256.
- **Transparency**: AWS KMS handles encryption/decryption automatically.

Encryption of Data in Transit

- **Purpose:** Secures data moving across networks.
- **Protocol:** Utilizes TLS (formerly SSL).
- **AWS Certificate Manager:** Manages and deploys TLS/SSL certificates.
- **Secure Connections:** HTTPS provides encrypted communication, preventing eavesdropping.
- **Examples:**
 - EC2 with Amazon EFS uses TLS for data traffic.
 - AWS Storage Gateway encrypts data when connecting to Amazon S3.

Securing Amazon S3 Buckets and Objects

- **Default Setting:** S3 buckets and objects are private by default.
- **Access Management:**
 - **Block Public Access:** Prevents unintended public access.
 - **IAM Policies:** Specify user access.
 - **Bucket Policies:** Define access at the bucket level.
 - **Access Control Lists (ACLs):** Legacy access control method.
 - **AWS Trusted Advisor:** Checks bucket permissions for security.

Working to ensure compliance:

AWS Compliance Programs

- **Overview:** AWS helps customers meet security and compliance requirements through various programs.
- **Categories:**
 - **Certifications and Attestations:** Assessed by independent auditors (e.g., ISO 27001, 27017, 27018).
 - **Laws and Regulations:** Compliance with legal frameworks like GDPR and HIPAA.
 - **Alignments and Frameworks:** Specific security requirements (e.g., CIS, EU-US Privacy Shield).
- **Resources:** Detailed information about AWS compliance programs and in-scope services can be found on the AWS compliance webpage.

AWS Config

- **Purpose:** Assesses, audits, and evaluates AWS resource configurations.
- **Features:**
 - Continuous monitoring of resource configurations.
 - Automates evaluations against desired configurations.
 - Tracks configuration changes and history.
- **Benefits:** Simplifies compliance auditing, security analysis, and operational troubleshooting.

AWS Artifact

- **Purpose:** Provides access to compliance-related documents.

- **Resources:** Offers downloadable security and compliance reports (e.g., ISO certifications, PCI reports).
- **Usage:** Helps customers demonstrate AWS service compliance and evaluate internal controls.
- **Agreements:** Allows review and acceptance of legal agreements like the Business Associate Agreement (BAA) for HIPAA compliance.

• **AWS security compliance programs** provide information about the policies, processes, and controls that are established and operated by AWS.

• **AWS Config** is used to assess, audit, and evaluate the configurations of AWS resources.

• **AWS Artifact** provides access to security and compliance reports.

Which of the following is AWS's responsibility under the AWS shared responsibility model?

| Choice | Response |
|----------|--|
| A | Configuring third-party applications |
| B | Maintaining physical hardware |
| C | Securing application access and data |
| D | Managing custom Amazon Machine Images (AMIs) |

The correct answer is B.

MODULE-5

Networking basics:

Networking Basics

- **Definition:** A computer network consists of two or more client machines connected to share resources.
- **Networking Devices:** Routers and switches are essential for connecting clients and enabling communication.

IP Addresses

- **IPv4 Addresses:**
 - Example: 192.0.2.0
 - Format: Four decimal numbers separated by dots, each representing 8 bits (0-255).
 - Total: 32 bits.
- **IPv6 Addresses:**
 - Example: 2600:1f18:22ba:8c00:ba86:a05e:a5ba:00FF
 - Format: Eight groups of four hexadecimal numbers separated by colons.
 - Total: 128 bits, accommodating more devices than IPv4.

Classless Inter-Domain Routing (CIDR)

- **Purpose:** CIDR allows for more efficient IP address allocation and routing.
- **Notation:** An IP address followed by a slash and a number (e.g., 192.0.2.0/24).
 - The number indicates how many bits are fixed for the network identifier.
 - Flexible bits determine the number of available IP addresses.
 - Example: 192.0.2.0/24 allows for 256 addresses (from 192.0.2.0 to 192.0.2.255).
- **Special Cases:**
 - Fixed IP Address: 192.0.2.0/32 for specific host access.
 - Entire Internet: 0.0.0.0/0 represents all possible IP addresses.

Open Systems Interconnection (OSI) Model

- **Overview:** The OSI model is a framework for understanding how data travels across a network, consisting of seven layers:
 1. **Physical Layer (1):** Transmission of raw bitstreams (signals).
 2. **Data Link Layer (2):** Data transfer within the same LAN (hubs and switches).
 3. **Network Layer (3):** Routing and packet forwarding (routers).
 4. **Transport Layer (4):** Host-to-host communication protocols (TCP, UDP).
 5. **Session Layer (5):** Orderly exchange of data (NetBIOS, RPC).
 6. **Presentation Layer (6):** Data readability and encryption.

7. **Application Layer (7):** Access methods for applications (HTTP, FTP, DHCP).

Amazon VPC:

Amazon VPC Overview

- **Purpose:** Enables provisioning of a logically isolated section of the AWS Cloud for launching AWS resources in a defined virtual network.
- **Control:** Users can control their networking resources, including:
 - Selection of IP address range
 - Creation of subnets
 - Configuration of route tables and network gateways
- **IP Addressing:** Supports both IPv4 and IPv6.

VPCs and Subnets

- **VPCs:**
 - Logically isolated and dedicated to a single AWS account.
 - Span a single AWS Region and can cover multiple Availability Zones.
- **Subnets:**
 - Ranges of IP addresses within a VPC, tied to a single Availability Zone.
 - Classified as **public** (direct internet access) or **private** (no internet access).

IP Addressing Details

- **CIDR Blocks:**
 - Assign a range of private IPv4 addresses upon VPC creation (e.g., /16 for 65,536 addresses, /28 for 16 addresses).
 - Subnet CIDR blocks cannot overlap.
- **Reserved IP Addresses:**
 - Five IP addresses are reserved for each subnet (e.g., network address, internal communications, DNS resolution).

Public IP Address Types

- **Public IPv4 Addresses:**
 - **Elastic IP Address:** Static and designed for dynamic use; associated with AWS accounts.
 - Automatically assigned through subnet settings or manually through Elastic IP.

Elastic Network Interface

- A virtual interface that can be attached/detached from instances, maintaining its attributes.
- Each instance has a default network interface with a private IPv4 address from the VPC range.

Route Tables

- Contains a set of rules directing network traffic from subnets.
- Each subnet must be associated with one route table.
- Includes a local route for communication within the VPC.

- A VPC is a logically isolated section of the AWS Cloud.
- A VPC belongs to one Region and requires a CIDR block.
- A VPC is subdivided into subnets.
- A subnet belongs to one Availability Zone and requires a CIDR block.
- Route tables control traffic for a subnet.
- Route tables have a built-in local route.
- You add additional routes to the table.
- The local route cannot be deleted.

VPC networking :

- **AWS Site-to-Site VPN:**
 - Allows connection between a VPC and an on-premises data center.
 - Involves creating a virtual private gateway and defining route tables for traffic routing.
- **AWS Direct Connect:**
 - Establishes a dedicated, private network connection between your data center and AWS.
 - Offers improved performance and reduced costs compared to internet-based connections.
- **VPC Endpoints:**
 - **Interface Endpoints:** Enable private connections to services powered by AWS PrivateLink.
 - **Gateway Endpoints:** Specifically for Amazon S3 and DynamoDB, offering no additional charge.
- **AWS Transit Gateway:**
 - Simplifies network management by acting as a hub that connects multiple VPCs and on-premises networks.
 - Reduces operational costs by centralizing connectivity management.
- **Network Performance Considerations:**
 - Importance of reducing latency by using direct connections for better bandwidth and performance.

- **VPC Networking Options:**

- **Internet Gateway:** Connects a VPC to the internet.
- **NAT Gateway:** Allows instances in private subnets to access the internet.
- **VPC Peering:** Connects multiple VPCs.
- **VPC Sharing:** Enables multiple accounts to share resources in a VPC.

- There are several VPC networking options, which include:
 - Internet gateway
 - NAT gateway
 - VPC endpoint
 - VPC peering
 - VPC sharing
 - AWS Site-to-Site VPN
 - AWS Direct Connect
 - AWS Transit Gateway
- You can use the VPC Wizard to implement your design.

VPC security :

Security Groups

- Control inbound/outbound traffic at the **instance level**.
- Default: deny all inbound, allow all outbound.
- **Stateful:** Return traffic is allowed automatically.

Network Access Control Lists (ACLs)

- Act at the **subnet level**.
- Default: allow all inbound/outbound traffic.
- **Stateless:** Must explicitly allow return traffic.

VPC Design Proposal

VPC CIDR Block: 10.0.0.0/16.

Public Subnet (Web Server):

- CIDR: 10.0.0.0/24.
- Route Table: Route to Internet Gateway.
- Security Group: Allow inbound on TCP ports 80 (HTTP) and 443 (HTTPS).

Private Subnet (Database Server):

- CIDR: 10.0.1.0/24.
- Route Table: Route to NAT Gateway.

- Security Group: Allow inbound from the web server on database port (e.g., TCP 3306).

NAT Gateway: Located in the public subnet for secure internet access for the database server.

High Availability: Deploy across multiple Availability Zones for redundancy.

- Build security into your VPC architecture:
 - Isolate subnets if possible.
 - Choose the appropriate gateway device or VPN connection for your needs.
 - Use firewalls.
- Security groups and network ACLs are firewall options that you can use to secure your VPC.

Amazon Route 53 :

Amazon Route 53 Overview

- **Type:** Highly available and scalable DNS web service.
- **Function:** Translates domain names (e.g., www.example.com) into IP addresses for routing user requests.
- **Compatibility:** Supports IPv4 and IPv6.
- **Integration:** Connects to AWS and external resources, monitoring health and performance.
- **Features:** Traffic management, health checks, and domain name registration.

Key Features

- **DNS Resolution:** Resolves domain queries and returns corresponding IP addresses.
- **Routing Policies:**
 - **Simple:** Basic round-robin routing.
 - **Weighted:** Traffic distribution based on assigned weights.
 - **Latency:** Routes to the lowest-latency AWS Region.
 - **Geolocation:** Routes based on user location.
 - **Failover:** Redirects to backup resources if primary fails.
 - **Multivalue Answer:** Returns multiple healthy DNS records.

Use Cases

- **Multi-Region Deployment:** Directs users to the closest AWS resources.
- **DNS Failover:** Enhances availability by redirecting traffic during outages using health checks.

- Amazon Route 53 is a highly available and scalable cloud DNS web service that translates domain names into numeric IP addresses.
- Amazon Route 53 supports several types of routing policies.
- Multi-Region deployment improves your application's performance for a global audience.
- You can use Amazon Route 53 failover to improve the availability of your applications.

Amazon CloudFront :

Content Delivery and Network Latency

When accessing websites, requests travel through various networks to reach the origin server, impacting performance and latency. A **Content Delivery Network (CDN)** addresses these issues.

What is a CDN?

- A globally distributed network of caching servers.
- Caches and delivers static content from nearby locations.
- Improves the speed of dynamic content delivery.

Amazon CloudFront

- A fast, secure CDN that delivers data globally with low latency.
- Utilizes **edge locations** for popular content and **Regional Edge Caches** for less popular content.
- Self-service model with pay-as-you-go pricing.

Benefits of Amazon CloudFront

1. **Fast and Global:** Low-latency content delivery.
2. **Security:** Built-in protections and SSL support.
3. **Customizable:** Programmable features for specific needs.
4. **AWS Integration:** Seamless use with other AWS services.
5. **Cost-Effective:** No minimum fees; charges based on usage.

Pricing

- **Data Transfer Out:** Charged per GB transferred.
- **HTTP(S) Requests:** Charged per request.

- **Invalidation Requests:** Free for the first 1,000 paths; charges apply beyond that.
- **Dedicated IP Custom SSL:** \$600/month per certificate.

- A CDN is a globally distributed system of caching servers that accelerates delivery of content.
- Amazon CloudFront is a fast CDN service that securely delivers data, videos, applications, and APIs over a global infrastructure with low latency and high transfer speeds.
- Amazon CloudFront offers many benefits.

Which AWS networking service enables a company to create a virtual network within AWS?

| Choice | Response |
|----------|--------------------|
| A | AWS Config |
| B | Amazon Route 53 |
| C | AWS Direct Connect |
| D | Amazon VPC |

The correct answer is D.

MODULE-6

Compute services overview:

AWS Compute Services Overview

AWS offers various compute services categorized into four main types:

Infrastructure as a Service (IaaS): Provides virtual machines for flexibility and control.

1. **Amazon EC2:** Launch and manage virtual machines with customizable configurations.

Serverless Computing: Run code without managing servers.

1. **AWS Lambda:** Execute code based on events or schedules, paying only for the compute time used.

Container-Based Computing: Efficiently run multiple workloads on a single OS.

1. **Amazon ECS:** Manage Docker containers.
2. **Amazon EKS:** Run Kubernetes for container orchestration.
3. **AWS Fargate:** Run containers without managing servers.
4. **Amazon ECR:** Store and manage Docker container images.

Platform as a Service (PaaS): Simplifies application deployment.

1. **AWS Elastic Beanstalk:** Deploy applications easily while AWS manages the underlying infrastructure.

| Key Features of Compute Services | | | |
|----------------------------------|-----------------|----------------------------------|--|
| Service | Category | Key Concepts | Ease of Use |
| Amazon EC2 | IaaS | Instance-based, Virtual Machines | Familiar for IT professionals |
| AWS Lambda | Serverless | Function-based, Low-cost | New concept, but user-friendly |
| Amazon ECS | Container-based | Instance-based | Quick job execution, admin overhead varies |
| AWS Fargate | Container-based | Reduces management overhead | Flexible control options |
| Elastic Beanstalk | PaaS | Web applications | Fast deployment, focus on code |

Amazon EC2 :

Amazon EC2 Overview:

Cost-Effective Solution: Running on-premises servers is expensive due to hardware procurement, data center maintenance, and provisioning for peak loads. EC2 offers a cloud-based alternative, allowing organizations to pay only for the resources they use.

Elastic Compute Cloud (EC2): EC2 provides virtual machines (instances) that can be quickly launched in various sizes and configurations, making it suitable for multiple workloads like application servers, web servers, and databases.

Key Features:

- **Elasticity:** Easily scale the number of servers and their sizes based on application needs.
- **Guest OS Control:** Full administrative control over Windows or Linux operating systems.
- **Rapid Deployment:** Launch instances in minutes from Amazon Machine Images (AMIs).

Launching an EC2 Instance:

- Use the **AWS Management Console Launch Instance Wizard** to simplify the process.
- Key decisions include selecting an AMI, instance type, and configuring network settings and security groups.

Amazon Machine Images (AMIs):

- AMIs serve as templates for launching EC2 instances, containing the operating system and installed applications.
- Different types of AMIs are available, including Quick Start, custom AMIs, and community AMIs.

Instance Types:

- EC2 offers various instance types optimized for different use cases (e.g., general-purpose, compute-optimized, memory-optimized).
- Instance names indicate the family, generation, and size, allowing for tailored resource allocation based on specific workloads.

Use Cases:

- Suitable for a variety of applications, including web hosting, game servers, media processing, and more.
- **Amazon EC2** enables you to run Windows and Linux **virtual machines** in the cloud.
- You launch **EC2 instances** from an **AMI** template into a VPC in your account.
- You can choose from many **instance types**. Each instance type offers different combinations of CPU, RAM, storage, and networking capabilities.
- You can configure **security groups** to control access to instances (specify allowed ports and source).
- **User data** enables you to specify a script to run the first time that an instance launches.
- Only **instances that are backed by Amazon EBS can be stopped**.
- You can use **Amazon CloudWatch** to capture and review metrics on EC2 instances.

Amazon EC2 cost optimization :

Amazon EC2 Pricing Models:

On-Demand Instances:

1. **Benefits:** Flexible, no long-term commitment, ideal for spiky or unpredictable workloads (e.g., development/testing).
2. **Use Cases:** Short-term applications, testing environments.

Spot Instances:

1. **Benefits:** Significant discounts, suitable for large-scale dynamic workloads that can tolerate interruptions.
2. **Use Cases:** Fault-tolerant applications, web servers, big data processing.

Reserved Instances:

1. **Benefits:** Predictable pricing for steady-state workloads, cost savings for long-term commitments.
2. **Use Cases:** Applications with consistent usage patterns.

Dedicated Hosts:

1. **Benefits:** Compliance with licensing restrictions, suitable for sensitive workloads.
2. **Use Cases:** Software requiring specific licensing or regulatory compliance.

Cost Optimization Strategies:

Right-Size

1. Choose the appropriate instance types based on workload needs.
2. Monitor CPU, RAM, and storage usage for potential downsizing.
3. Use Amazon CloudWatch for metrics and insights.

Increase Elasticity:

1. Utilize automatic scaling to manage peak loads effectively.
2. Turn off non-production instances during non-business hours.

Optimal Pricing Model:

1. Analyze workload patterns to select the best pricing model.
2. Combine On-Demand, Spot, and Reserved Instances to match usage needs.

Optimize Storage Choices:

1. Resize EBS volumes and select cost-effective storage options.
2. Delete unnecessary EBS snapshots and implement data lifecycle policies.

- **Amazon EC2 pricing models** include On-Demand Instances, Reserved Instances, Spot Instances, Dedicated Instances, and Dedicated Hosts.
- **Spot Instances** can be interrupted with a 2-minute notification. However, they can offer significant cost savings over On-Demand Instances.
- The **four pillars of cost optimization** are:
 - Right size
 - Increase elasticity
 - Optimal pricing model
 - Optimize storage choices

Container services :

Docker Overview

- **What is Docker?:** A platform to build, test, and deploy applications using containers, which are lightweight packages that include everything needed to run an application.
- **Benefits:** Standardizes environments, reduces version conflicts, facilitates microservices, and enhances portability.

Containers vs. Virtual Machines (VMs)

- **Containers:** Share the host OS kernel, making them lightweight and efficient.
- **VMs:** Require separate OS instances, leading to higher resource usage.

AWS Services

- **Amazon ECS:** A scalable service for managing Docker containers, integrates with AWS services, and uses task definitions to define container specifications.
- **Amazon EKS:** A managed Kubernetes service for running containerized applications, certified for compatibility with Kubernetes.
- **Amazon ECR:** A fully managed Docker container registry for storing and deploying container images, integrated with ECS and EKS.

- **Containers** can hold everything that an application needs to run.
- **Docker** is a software platform that packages software into containers.
 - A single application can span multiple containers.
- Amazon Elastic Container Service (**Amazon ECS**) orchestrates the running of Docker containers.
- **Kubernetes** is open source software for container orchestration.
- Amazon Elastic Kubernetes Service (**Amazon EKS**) enables you to run Kubernetes on AWS
- Amazon Elastic Container Registry (**Amazon ECR**) enables you to store, manage, and deploy your Docker containers.

Introduction to AWS Lambda :

Overview of AWS Lambda

- **Programming Language Support:** Supports languages like Java, Go, PowerShell, Node.js, C#, Python, and Ruby.
- **Automation:** Fully automated administration, including maintenance and security patches.
- **Fault Tolerance:** Built-in fault tolerance with compute capacity across multiple Availability Zones.
- **Cost-Effective:** Pay only for the requests and compute time used, billed in 100 ms increments.

Event Sources

- **Definition:** AWS services or applications that trigger Lambda functions.
- **Examples:** Amazon S3, Amazon SNS, Amazon CloudWatch Events, Amazon SQS, DynamoDB, Application Load Balancer, and API Gateway.

Function Configuration

- **Setup:** Specify runtime environment, execution role, triggers, memory allocation, and other configurations.
- **Deployment Package:** Code and dependencies packaged into a ZIP archive.

Use Case Examples

- **Schedule-Based:** Automate starting/stopping EC2 instances using CloudWatch Events.
- **Event-Based:** Create thumbnails for images uploaded to S3 using Lambda.

Quotas and Limits

- **Resource Limits:** Max memory allocation is 10,240 MB, max run time is 15 minutes, and deployment package size is 250 MB (or 10 GB for container images).
- **Concurrent Executions:** Limited to 1,000 per Region.

Key Takeaways

- **Serverless Computing:** Build applications without server management.
- **Event Sources Trigger Functions:** Functions are triggered by AWS events.
- **Automatic Scaling and Fault Tolerance:** AWS Lambda handles scaling automatically

Introduction to AWS Elastic Beanstalk :

Overview of AWS Elastic Beanstalk

- **Purpose:** Simplifies the deployment of web applications by managing infrastructure tasks.
- **Managed Service:** Handles:
 - Infrastructure provisioning and configuration
 - Deployment
 - Load balancing
 - Automatic scaling
 - Health monitoring
 - Analysis and debugging
 - Logging
- **Cost:** No additional charges; pay only for the underlying resources used.

Deployment Options

- **Supported Platforms:** Docker, Go, Java, .NET, Node.js, PHP, Python, Ruby.
- **Deployment Methods:** Use the AWS Management Console, AWS CLI, Visual Studio, or Eclipse to upload your application.
- **Application Servers:** Utilizes servers like Apache Tomcat, NGINX, Apache HTTP Server, Passenger, Puma, and Microsoft IIS.

Key Features

- **Ease of Use:** Quick to set up and deploy applications without worrying about server management.
- **Automatic Scaling:** Adjusts resources based on application needs and traffic peaks, improving cost efficiency.
- **Developer Focus:** Increases productivity by allowing developers to concentrate on writing code instead of managing infrastructure.

Benefits of Elastic Beanstalk

- **Fast Setup:** Easy to get started with minimal configuration.
- **Increased Productivity:** Focus on coding rather than infrastructure management.
- **Scalability:** Can handle growing application demands.
- **Control:** Full control over AWS resources used for the application.

- **AWS Elastic Beanstalk** enhances developer productivity.
 - Simplifies the process of deploying your application.
 - Reduces management complexity.
- Elastic Beanstalk supports **Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker**
- There is no charge for Elastic Beanstalk. Pay only for the AWS resources that you use.

Which AWS service helps developers quickly deploy resources which can make use of different programming languages, such as .NET and Java?

| Choice | Response |
|----------|---|
| A | AWS CloudFormation |
| B | AWS SQS |
| C | AWS Elastic Beanstalk |
| D | Amazon Elastic Compute Cloud (Amazon EC2) |

The correct answer is C.

Section1-EBS

Amazon Elastic Block Store (EBS) provides reliable, block-level storage for Amazon EC2 instances. It functions like a virtual hard drive that you can attach to your EC2 instances, with data automatically replicated within the same Availability Zone to ensure durability.

Key Features:

1. **Storage Types:**
 - **SSD (Solid State Drives):**
 - **General Purpose SSD:** Best for everyday workloads.
 - **Provisioned IOPS SSD:** Designed for high-performance, I/O-intensive applications (e.g., large databases).
 - **HDD (Hard Disk Drives):**
 - **Throughput-Optimized HDD:** Good for streaming workloads requiring consistent, fast throughput.
 - **Cold HDD:** Ideal for infrequently accessed data with lower storage costs.
2. **Snapshots:**
 - Backups of your EBS volumes stored in Amazon S3.
 - Snapshots are incremental, meaning only changes from the last snapshot are saved.
 - Can be used to restore data or create new volumes at any time.
3. **Elasticity:**
 - You can resize volumes (increase capacity) or change types (e.g., from HDD to SSD) without stopping your EC2 instance.
4. **Encryption:**
 - Data moving between your EC2 instance and EBS is encrypted by default, with no extra cost.

Use Cases:

- Boot volumes for EC2 instances.
- Storage for databases and enterprise applications.
- High-throughput applications such as big data analytics.

Pricing:

- **Volume Size:** Charged based on the size you provision (per GB/month).
- **IOPS:** Charged based on input/output operations per second for Provisioned IOPS SSDs.
- **Snapshots:** Additional cost per GB of data stored in S3.

- **Data Transfer:** Free for inbound data; charges apply for outbound transfers between regions.

EBS offers flexibility, performance, and security, making it suitable for a wide range of workloads, from basic storage to high-performance databases.

Section 2 S3

Amazon S3 is a cloud storage service by AWS that automatically scales with your data and application needs. Here's a clear breakdown:

Key Features:

1. **Automatic Scaling:** S3 handles storage growth and high request volumes automatically, so you don't need to manage storage capacity.
2. **Multiple Access Methods:** You can access S3 through:
 - AWS Management Console
 - AWS CLI (Command Line Interface)
 - AWS SDK (Software Development Kit)
 - REST-based URLs (HTTP/HTTPS)
3. **Global Access:** S3 bucket names must be globally unique, and you can access data anywhere via a secure URL.

Common Use Cases:

- **Application Storage:** S3 provides a shared space for storing application data like user media files and logs, accessible from any instance, such as Amazon EC2.
- **Static Website Hosting:** Host static content like HTML, CSS, and JavaScript directly from S3.
- **Backup & Disaster Recovery:** Store backups with S3's high durability (99.999999999%) and enable cross-region replication for added data security.
- **Media Hosting:** Store and manage videos, photos, or music files with scalability.
- **Software Distribution:** Host software files for customer downloads.

Pricing:

- **Pay for What You Use:**
 - Based on the storage size (GB per month).
 - Charges apply for different request types like PUT, GET, COPY, etc.
 - **Free:** Transfers into S3 and transfers to services like CloudFront or EC2 within the same region.
 - **Charges:** Data transfers out of the region incur costs.

Storage Classes:

- **S3 Standard:** For frequently accessed data, offering 99.999999999% durability and 99.99% availability.
- **S3 Infrequent Access (S-IA):** For data that's accessed less often, providing lower costs but slightly less availability (99.9%).

Summary:

Amazon S3 provides a flexible, scalable, and cost-effective solution for storing and accessing data, with robust security and global accessibility. You only pay for what you use, making it ideal for a wide range of applications, from backups to large-scale media hosting.

Section 3: Amazon Elastic File System (Amazon EFS)

Amazon Elastic File System (EFS) provides scalable, elastic, and fully managed file storage for use with AWS services and on-premises resources. It allows multiple Amazon EC2 instances to access a shared file system simultaneously using the Network File System (NFS) protocol.

Key Features of Amazon EFS:

1. **Simple, Elastic Storage:** Automatically scales based on file usage, growing or shrinking as needed.
2. **Shared File System:** Accessible by multiple EC2 instances across Availability Zones, offering consistent performance and file locking.
3. **NFS Support:** Compatible with NFS versions 4.0 and 4.1, and works seamlessly with Linux-based EC2 instances.

4. **Wide Range of Use Cases:** Ideal for big data analytics, media workflows, content management, web serving, and home directories.
5. **Petabyte-Scale Storage:** Automatically scales from gigabytes to petabytes without provisioning.

Architecture:

EFS operates within a Virtual Private Cloud (VPC) and is designed to support access from multiple Availability Zones (AZs). For optimal performance, it is recommended to access the file system from a mount target within the same AZ.

Setup Process:

1. **Launch EC2 Resources:** Create EC2 instances in your VPC.
2. **Create EFS File System:** Set up your EFS file system and configure mount targets.
3. **Mount File System:** Attach the EFS file system to your EC2 instance.
4. **Test and Verify:** Verify that the setup works as expected.
5. **Clean Up:** Remove unused resources and secure your account.

Resources:

- **File System:** Core component with properties like size, ID, and mount targets.
- **Mount Targets:** Connect EC2 instances to the file system within specific VPC subnets.
- **Tags:** Metadata that can be used to organize and manage file systems.

Key Takeaways:

- Amazon EFS is designed for flexible, scalable network file storage.
- Suitable for a variety of workloads, from analytics to media processing.
- It is fully managed, eliminating storage administration tasks.
- Available through the AWS Console, API, or CLI, and you pay only for what you use.

Section 4 Amazon S3 Glacier

Amazon S3 Glacier is a highly secure, durable, and cost-effective cloud storage service designed for long-term data archiving. While offering extremely low storage costs, retrieval times are longer compared to Amazon S3, making it ideal for infrequently accessed data.

Key Components:

- **Archive:** The basic unit of storage (e.g., a file, photo, or video) with a unique ID.
- **Vault:** A container for storing archives. Each vault is region-specific and has its own access policy.
- **Vault Access Policy:** Defines who can access the vault and what operations can be performed. A vault lock policy ensures compliance by preventing unauthorized alterations.

Data Retrieval Options:

1. **Expedited Retrievals:** Available in 1-5 minutes at the highest cost.
2. **Standard Retrievals:** Available in 3-5 hours.
3. **Bulk Retrievals:** Available in 5-12 hours at the lowest cost.

Key Features:

- **Low-Cost Storage:** S3 Glacier is one of the most cost-efficient solutions for long-term storage.
- **Durability:** Provides 11 nines (99.999999999%) of durability.
- **Encryption:** Data is encrypted in transit and at rest by default with AES-256.
- **Compliance:** Vault Lock enforces compliance, ideal for regulatory needs like healthcare and financial services.

Use Cases:

- **Media Archiving:** Ideal for storing large volumes of media content.
- **Healthcare:** Stores electronic health records and complies with regulations.
- **Regulatory Archives:** For financial and compliance purposes.
- **Scientific Data Archiving:** Secure long-term storage for research data.
- **Digital Preservation:** Used by libraries and government agencies.

Lifecycle Policies:

Amazon S3 allows automatic data transition from S3 Standard to S3 Glacier as part of a lifecycle policy, optimizing cost by archiving older, less accessed data.

Comparison with Amazon S3:

- **Latency:** S3 Glacier has longer retrieval times compared to the low-latency Amazon S3.
- **Item Size:** S3 Glacier supports up to 40 TB, whereas S3 is limited to 5 TB per object.
- **Cost:** S3 Glacier offers lower storage costs but higher retrieval costs.

Conclusion:

Amazon S3 Glacier is perfect for organizations that require long-term, durable storage at low costs with infrequent access to data, making it a robust solution for data archiving and compliance.

Module check

1. True or False? Amazon Simple Storage Service (Amazon S3) is an object storage suitable for the storage of flat files like Microsoft Word documents, photos, etc.

☒ True

☐ False

2. Amazon S3 replicates all objects _____. (Select the best answer)

☐ on multiple volumes within an Availability Zone

☒ in multiple Availability Zones within the same Region

☐ across multiple Regions for higher durability

☐ on multiple S3 buckets

3. Which of the following can be used as a storage class for an S3 object lifecycle policy? (Choose three)

☒ S3 - Standard Access

☐ AWS Storage Gateway

☒ S3 - Infrequent Access

☒ Simple Storage Service Glacier

☐ S3 - Reduced Redundancy Storage

☐ Amazon Dynamo DB

4. The name of an S3 bucket must be unique _____. (Select the best answer)

- ☒ worldwide across all AWS accounts
- ☐ within a Region
- ☐ across all your AWS accounts
- ☐ within your AWS account

5. You can use Amazon Elastic File System (Amazon EFS) to: (Select the best answer)

- ☐ provide simple, scalable, elastic file storage for use only within AWS.
- ☒ implement storage for Amazon EC2 instances that multiple virtual machines can access at the same time.
- ☐ host a robust CDN to deliver entire web sites with dynamic, static, and streaming content.
- ☐ generate user-specific content.

6. Amazon Elastic Block Store (Amazon EBS) is recommended when data _____ and _____. (Choose two)

- ☐ requires object-level storage
- ☒ must be quickly accessible, requiring long-term persistence
- ☒ requires an encryption solution
- ☐ needs to be stored in a different Availability Zone than the one the EC2 instance is in

7. True or False? By default, all data stored in Amazon S3 is viewable by the public.

- ☐ True
- ☒ False

8. Regarding Amazon S3 Glacier, what is a Vault? (Select the best answer)

- ☐ The rules that determine who may (or may not) access archives
- ☐ An object (photos, videos, files, or documents)
- ☒ A container for storing archives
- ☐ A policy that identifies who can access content stored in Glacier

9. True or False? When you create a bucket in Amazon S3, it is associated with a specific AWS Region.

- ☒ True
- ☐ False

10. Which of the following are features of Amazon Elastic Block Store (Amazon EBS)? (Choose two)

- ☒ Data stored on Amazon EBS is automatically replicated within an Availability Zone.
- ☐ Amazon EBS data is automatically backed up to tape.
- ☐ Data on an Amazon EBS volume is lost when the attached instance is stopped.
- ☒ Amazon EBS volumes can be encrypted transparently to workloads on the attached instance.

Section 1 Amazon RDS

(Relational Database Service) features, deployment options, and pricing models:

1. High Availability with Multi-AZ Deployment:

- Amazon RDS ensures high availability by automatically failing over to a standby instance in a different Availability Zone. Synchronous replication minimizes potential data loss, and no changes are required in application code for failover.

2. Read Replicas:

- Used for read-heavy workloads, Amazon RDS read replicas offer asynchronous replication. They can offload read queries and be promoted to the primary instance if needed. They can also be located in different regions to reduce latency or serve disaster recovery needs.

3. Use Cases:

- RDS works well for web and mobile applications, e-commerce platforms, and online games that require high throughput, scalability, and availability. It provides a secure, low-cost, fully managed database solution.

4. When to Use RDS:

- Ideal for complex transactions and queries, with moderate to high IOPS (up to 30,000). RDS is not suitable for massive write rates, sharding, or simple queries best handled by NoSQL databases.

5. Pricing and Deployment Considerations:

- Amazon RDS offers on-demand and reserved instance options, with the latter offering significant cost savings for long-term usage. Backup storage for active databases is included, while additional and backup storage for terminated instances is billed per GB.

6. Storage and Data Transfer:

- Storage costs vary based on deployment type (Single vs. Multi-AZ). Inbound data transfer is free, while outbound transfer costs are tiered.

Section 2: Amazon DynamoDB - Summary

1. DynamoDB Overview:

- DynamoDB is a fast, scalable, and flexible **NoSQL database** service offering single-digit millisecond latency at any scale.
- Supports **key-value** and **document store** models with **virtually unlimited storage**.

2. Core Components:

- **Tables, items, and attributes**: Data is organized in tables, where items represent rows, and attributes represent columns.
- **Primary Keys**: Supports **partition keys** (simple key) or **composite keys** (partition and sort keys).

3. Key Features:

- **High scalability**: Handles both read and write throughput, either manually or via **auto-scaling**.
- **Global tables**: Automatically replicate data across AWS regions.
- **Encryption at rest** and **Time-to-Live (TTL)** features to ensure data security and management.

4. Performance:

- Runs on **SSDs** for low-latency queries.
- Enables both **query** operations using the primary key and **scan** operations for non-key attributes (though scans are less efficient).

5. Use Cases:

- Ideal for **mobile, web, gaming, adtech, and IoT** applications where low-latency and high-scale performance are critical.
- Works well for systems requiring **horizontal scaling** of large amounts of semi-structured and unstructured data.

6. Access:

- Accessible via the **AWS console**, **CLI**, and **API calls**.

DynamoDB is designed for applications requiring scalable storage and consistent low-latency performance.

Section 3: Amazon Redshift - Summary

1. Overview of Amazon Redshift:

- **Amazon Redshift** is a fast, fully managed data warehouse designed for simple and cost-effective data analysis using standard SQL and existing business intelligence (BI) tools.
- It allows users to run complex analytical queries against petabytes of structured data quickly and efficiently.

2. Key Features:

- **Parallel Processing Architecture:** Utilizes a leader node for managing communications and compute nodes for processing queries, resulting in rapid query responses.
- **Cost Efficiency:** Start for as little as 25 cents per hour, with scalable storage and processing costs around \$1,000 per terabyte per year.
- **Amazon Redshift Spectrum:** Enables querying of large datasets directly in Amazon S3.

3. Automation and Scalability:

- Simplifies the management, monitoring, and scaling of Redshift clusters.
- Clusters can be scaled up or down with ease, adapting to changing needs without downtime.

4. Security and Compatibility:

- Built-in security with strong encryption for data both at rest and in transit.
- Compatible with existing SQL clients and BI tools through JDBC and ODBC connectors.

5. Use Cases:

- **Enterprise Data Warehouse (EDW):** Customers migrate traditional data warehouses for agility and reduced upfront costs.
- **Big Data:** Suitable for handling massive datasets without stretching existing systems.
- **Software as a Service (SaaS):** Scales capacity as demand grows, adding analytical functionality to applications.

6. Key Takeaways:

- Amazon Redshift is a **fully managed** data warehouse service that offers **fast performance**, **automatic scaling**, and **built-in encryption**.
- It leverages **columnar storage** and **parallel processing** to optimize data and query performance.
- Automatic monitoring and backup features ensure data safety and easy restoration when needed.

In summary, Amazon Redshift provides a robust, scalable solution for organizations looking to efficiently manage and analyze large volumes of data with minimal administrative overhead

Section 4: Amazon Aurora - Summary

1. Overview of Amazon Aurora:

- **Amazon Aurora** is an enterprise-class relational database service compatible with MySQL and PostgreSQL.
- It combines the performance and availability of high-end commercial databases with the simplicity and cost-effectiveness of open-source solutions, aiming to reduce costs while enhancing reliability and availability.

2. Key Benefits:

- **High Availability:** Aurora stores multiple copies of data across multiple Availability Zones and provides continuous backups to Amazon S3, ensuring data durability.
- **Performance and Scalability:** Features a fast, distributed storage subsystem and is straightforward to set up with standard SQL queries.
- **Managed Service:** Automates time-consuming tasks such as provisioning, patching, backup, recovery, failure detection, and repair.
- **Pay-as-you-go Pricing:** Customers pay only for the services and features they use.

3. High Availability and Resilience:

- Designed for high availability with the ability to use up to 15 read replicas to prevent data loss.
- Instant crash recovery is a core feature, allowing recovery in under 60 seconds after a crash without needing to replay redo logs.

4. Resilient Design:

- The buffer cache is separated from the database process, allowing immediate availability upon restart and reducing the impact on access during cache repopulation.

5. Security Features:

- Provides multiple levels of security, including:
 - **Network Isolation** using Amazon VPC.
 - **Encryption at Rest** with AWS Key Management Service (KMS).
 - **Encryption of Data in Transit** using Secure Sockets Layer (SSL).

6. Compatibility:

- Fully compatible with existing MySQL and PostgreSQL databases, allowing users to leverage existing tools with minimal changes.

7. Key Takeaways:

- Amazon Aurora is a **highly available, performant, and cost-effective** managed relational database solution.
- Its fault-tolerant, self-healing storage system is designed for the cloud, continuously backing up data and replicating it across multiple Availability Zones.
- Aurora automates many database management tasks, freeing users to focus on application development and data analysis.

In summary, Amazon Aurora offers a robust, efficient, and secure database solution, making it an excellent choice for businesses looking for high availability and performance without the complexities of traditional database management.

| What are my requirements? | |
|---|-------------------------------------|
| Enterprise-class relational database | Amazon RDS |
| Fast and flexible NoSQL database service for any scale | Amazon DynamoDB |
| Operating system access or application features that are not supported by AWS database services | Databases on Amazon EC2 |
| Specific case-driven requirements (machine learning, data warehouse, graphs) | AWS purpose-built database services |

Module check

1. You are designing an ecommerce web application that will scale to hundreds of thousands of concurrent users. Which database technology is best suited to hold the session state in this example?
 - ☐ Amazon Relational Database Service (Amazon RDS)
 - ☒ Amazon DynamoDB
 - ☐ Amazon Redshift
 - ☐ Amazon Simple Storage Service (Amazon S3)

2. You need to find an item in an Amazon DynamoDB table using an attribute other than the item's primary key. Which of the following operations should you use? (Select the best answer.)
- ☐ PutItem
 - ☒ Scan
 - ☐ Query
 - ☐ GetItem
3. In Amazon DynamoDB, what does the query operation enable you to do? (Select the best answer.)
- ☐ Query a table using the partition key and an optional sort key filter
 - ☐ Query any secondary indexes that exist for a table
 - ☐ Efficiently retrieve items from a table or secondary index
 - ☒ All of the above
4. Which AWS Cloud service is best suited for analyzing your data by using standard structured query language (SQL) and your existing business intelligence (BI) tools? (Select the best answer.)
- ☐ Amazon Relational Database Service (Amazon RDS)
 - ☐ Amazon Simple Storage Service Glacier
 - ☐ Amazon DynamoDB
 - ☒ Amazon Redshift
5. In Amazon DynamoDB, an attribute is _____. (Select the best answer.)
- ☒ a fundamental data element
 - ☐ a collection of items
 - ☐ a collection of attributes

6. If you are developing an application that requires a database with extremely fast performance, fast scalability, and flexibility in the database schema, which service should you consider? (Select the best answer.)
- ☐ Amazon Relational Database Service (Amazon RDS)
 - ☐ Amazon ElastiCache
 - ☒ Amazon DynamoDB
 - ☐ Amazon Redshift
7. Which of the following use cases is appropriate for using Amazon Relational Database Service (Amazon RDS)? (Select the best answer.)
- ☐ Massive read/write rates
 - ☐ Simple GET or PUT requests
 - ☒ Complex transactions
 - ☐ All of the above
8. A company has an application, which consists of a .NET layer that connects to a MySQL database. They want to move this application on to AWS and use AWS features such as high availability and automated backups. Which of the following would be an ideal database for this use case? (Select the best answer)
- ☐ Amazon Redshift
 - ☐ Amazon DynamoDB
 - ☐ Amazon RDS
 - ☒ Amazon Aurora

9. True or false? Amazon RDS automatically patches the database software and backs up your database, storing the backups for a user-defined retention period and enabling point-in-time recovery.

☒ True

☐ False

10. What should you consider when choosing a database type? (Select the best answer.)

☐ Data size

☐ Data access period

☐ Query frequency

☐ Highly available

☒ All of the above

Section 1 AWS Well-Architected Framework Pillars

The AWS Well-Architected Framework consists of five pillars that provide a structured approach to evaluating and improving cloud architectures. Each pillar addresses specific aspects of architecture design, helping organizations build secure, high-performing, resilient, and efficient infrastructure for applications in the cloud.

1. Operational Excellence

- **Focus:** Improve the ability to run and monitor systems, gain insight into operations, and continuously improve processes and procedures.
- **Key Topics:**
 - **Monitoring:** Implement effective monitoring and alerting mechanisms.
 - **Incident Response:** Develop incident response plans and conduct regular drills.
 - **Change Management:** Establish processes for managing changes to the system.
 - **Automation:** Use automation to manage tasks and workflows efficiently.

2. Security

- **Focus:** Protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies.
- **Key Topics:**
 - **Identity and Access Management:** Control access to resources using strong authentication and authorization practices.
 - **Data Protection:** Implement encryption for data at rest and in transit.
 - **Infrastructure Protection:** Use network segmentation, firewalls, and security groups to safeguard systems.
 - **Incident Response:** Prepare for security incidents with a defined response plan.

3. Reliability

- **Focus:** Ensure a workload performs its intended function correctly and consistently when it's expected to.
- **Key Topics:**
 - **Foundations:** Design applications with high availability and fault tolerance.
 - **Change Management:** Implement processes to manage updates without impacting availability.
 - **Backup and Recovery:** Establish backup strategies and recovery plans to minimize downtime.
 - **Testing and Monitoring:** Continuously test and monitor for reliability.

4. Performance Efficiency

- **Focus:** Use computing resources efficiently to meet system requirements and to maintain performance as demand changes.
- **Key Topics:**

- **Selection:** Choose the right resource types and sizes based on workloads.
- **Monitoring:** Continuously monitor resource utilization and performance metrics.
- **Tradeoffs:** Understand tradeoffs between cost and performance and apply optimizations as needed.
- **Evolving Workloads:** Adapt architectures to take advantage of new technologies and services.

5. Cost Optimization

- **Focus:** Avoid unnecessary costs while delivering business value through efficient resource management.
- **Key Topics:**
 - **Cloud Financial Management:** Implement practices for managing cloud expenditures.
 - **Cost-Effective Resources:** Choose the right type and size of resources to meet performance targets within budget.
 - **Usage Monitoring:** Track resource usage and spending to identify waste.
 - **Scaling:** Scale resources dynamically to match demand without overspending.

Summary

Each of these five pillars provides best practices and considerations that organizations can use to assess their cloud architectures. By evaluating workloads against these pillars, organizations can ensure that their cloud environments are well-architected, which leads to improved performance, security, reliability, and cost efficiency.

Section 2 Summary: Reliability and Availability

In this section, we explored the concepts of reliability and availability in the context of cloud architecture, emphasizing the need to plan for failures and downtime, as highlighted by Werner Vogels, CTO of Amazon.

Key Concepts:

1. Reliability

- **Definition:** Reliability measures a system's ability to provide desired functionality when required by the user.
- **Components:** It encompasses all components of the system (hardware, firmware, software).

- **Measurement:** Commonly measured by **Mean Time Between Failures (MTBF)**, which is calculated as the total time in service divided by the number of failures.
 - **Example:** If an application runs for 96 hours before failing and takes 72 hours to repair, the MTBF would be 168 hours.
2. **Availability**
- **Definition:** Availability is the percentage of time a system operates normally, calculated as normal operation time divided by total time.
 - **Expression:** It's often expressed in terms of "nines," such as five 9s (99.999% availability).
 - **Implications:** Any time the system is not operating as expected (scheduled or unscheduled downtime) reduces availability.
3. **High Availability**
- **Characteristics:** A highly available system can tolerate some level of degradation while remaining accessible, minimizing downtime and requiring minimal human intervention.
 - **Design Considerations:** It often involves shared resources and rapid restoration of services after failures.
4. **Availability Tiers**
- **Design Goals:** Different applications have varying availability requirements, defined by acceptable disruption lengths per year.
 - **Example:** A system designed for 99.9% availability can tolerate certain amounts of downtime, which varies depending on application criticality.
5. **Factors Influencing Availability**
- **Fault Tolerance:** The ability of an application to remain operational despite component failures, usually supported by redundant hardware.
 - **Scalability:** The capability to handle increasing capacity needs without significant changes to the design.
 - **Recoverability:** The ability to restore services quickly after a failure, ensuring minimal data loss.

Key Takeaways:

- **Reliability** is crucial for ensuring systems function when needed, while **availability** focuses on the percentage of uptime.
- Factors such as fault tolerance, scalability, and recoverability play significant roles in determining overall availability.
- Designing for high availability typically incurs additional costs, necessitating a balance between cost and user experience.

This section emphasizes that effective cloud architecture must incorporate these reliability and availability principles to build resilient systems capable of meeting user demands in the face of inevitable failures.

Section 3 Summary: AWS Trusted Advisor

In this section, we learned about **AWS Trusted Advisor**, a valuable tool designed to assist users in optimizing their AWS environments according to best practices. Trusted Advisor provides real-time guidance across various categories, enabling users to enhance performance, reduce costs, improve security, and ensure fault tolerance.

Key Features of AWS Trusted Advisor:

1. **Real-Time Guidance:**
 - AWS Trusted Advisor analyzes your entire AWS environment and offers actionable recommendations to help you provision resources effectively.
2. **Categories of Recommendations:** Trusted Advisor provides insights in five key categories:
 - **Cost Optimization:**
 - Identifies unused and idle resources, offering suggestions to optimize costs. This may include making commitments to reserved capacity to save on expenditures.
 - **Performance:**
 - Focuses on enhancing the performance of your services by monitoring service limits, ensuring efficient use of provisioned throughput, and identifying overutilized instances.
 - **Security:**
 - Aims to strengthen application security by identifying potential vulnerabilities, recommending AWS security features, and reviewing user permissions to close security gaps.
 - **Fault Tolerance:**
 - Provides suggestions to improve the availability and redundancy of applications, such as implementing automatic scaling, health checks, Multi-AZ deployments, and backup strategies.
 - **Service Limits:**
 - Monitors usage of AWS services and alerts you when usage exceeds 80% of service limits, helping prevent disruptions caused by reaching resource limits.
3. **Snapshot-Based Analysis:**
 - Trusted Advisor's recommendations are based on snapshots of your environment, which may take up to 24 hours to reflect any changes in usage or limits.

Conclusion:

AWS Trusted Advisor is an essential tool for AWS users, facilitating the implementation of best practices in resource provisioning. By leveraging its recommendations, users can optimize costs, enhance performance, strengthen security, and improve fault tolerance in their cloud environments. This ultimately leads to more efficient and reliable architectures on the AWS platform.

AWS Trusted Advisor

is an online tool that provides real time guidance to help you provision your resources by following AWS best practices.

-

AWS Trusted Advisor looks at your entire AWS environment and gives you real time recommendations in five categories.

-

You can use AWS Trusted Advisor to help you optimize your AWS environment as soon as you start implementing your architecture designs.

Module check

1. Which of the following is not one of the four areas of the performance efficiency pillar of the AWS Well-Architected Framework? (Select the best answer.)

☐ Tradeoffs

☐ Selection

☒ Traceability

☐ Monitoring

2. Which of the following is a principle when designing cloud-based systems? (Select the best answer.)

- ☐ Build tightly-coupled components
- ☐ Make infrequent, large batch changes
- ☒ Assume everything will fail
- ☐ Use as many services as possible

3. Which of the following are pillars of the AWS Well-Architected Framework? (Choose three.)

- ☒ Security
- ☐ Persistence
- ☒ Operational Excellence
- ☒ Cost Optimization

4. Which design principles are recommended when considering performance efficiency?
(Choose two.)

- ☐ Match supply with demand
- ☐ Analyze and attribute expenditure
- ☒ Democratize advanced technologies
- ☒ Use serverless architectures
- ☐ Enable traceability

5. AWS Trusted Advisor provides insight regarding which five categories of an AWS account? (Select the best answer.)

- ☐ Performance, cost optimization, access control, connectivity, security
- ☒ Performance, cost optimization, security, fault tolerance, service limits
- ☐ Security, fault tolerance, high availability, connectivity, service limits
- ☐ Security, access control, high availability, performance, service limits

6. What is the focus of the sustainability pillar of the Well-Architected Framework? (Select the best answer.)

- ☐ Designing workloads that recover quickly from failures
- ☒ Minimizing the environmental impacts of running cloud workloads
- ☐ Avoiding unnecessary costs in cloud workloads
- ☐ Automating updates to cloud workloads

7. After you move to the AWS Cloud, you want to ensure that the right security settings are put in place. Which online tool can assist in security compliance? (Select the best answer.)

- ☐ Amazon Kinesis
- ☐ AWS Support
- ☒ AWS Trusted Advisor
- ☐ Amazon CloudWatch

8. Which of the following is a measure of your system's ability to provide functionality when desired by the user? (Select the best answer.)

- ☐ Availability
- ☐ Fault Tolerance
- ☒ Reliability
- ☐ Performance efficiency

9. What is defined as the ability for a system to remain operational even if some of the components of that system fail? (Select the best answer.)

- ☐ High durability
- ☒ Fault tolerance
- ☐ High availability
- ☐ High durability

10. Which of the following best describes a system that can withstand some measures of degradation, experiences minimal downtime, and requires minimal human intervention? (Select the best answer.)

- ☐ Scalable
- ☐ Fault-tolerant
- ☐ Elastic
- ☒ Highly available

Section 1 Summary: Elastic Load Balancing

In this section, we explored **Elastic Load Balancing (ELB)**, which is a key service in AWS that helps distribute incoming application or network traffic across multiple targets in one or more Availability Zones, ensuring high availability and fault tolerance.

Types of Load Balancers:

ELB supports three types of load balancers, each suited for different use cases:

1. **Application Load Balancer (ALB):**
 - Operates at the application level (OSI model layer 7).
 - Routes traffic based on request content (ideal for HTTP/HTTPS traffic).
 - Supports advanced request routing and is designed for modern application architectures like microservices.
 - Ensures security by using the latest SSL/TLS protocols.
2. **Network Load Balancer (NLB):**
 - Operates at the network transport level (OSI model layer 4).
 - Routes connections based on IP protocol data and handles TCP and UDP traffic.
 - Capable of managing millions of requests per second with ultra-low latency.
 - Optimized for sudden and volatile traffic patterns.
3. **Classic Load Balancer (CLB):**
 - Provides basic load balancing across multiple EC2 instances at both application and transport levels.
 - Supports HTTP, HTTPS, TCP, and SSL but is considered an older implementation.
 - AWS recommends using ALB or NLB when possible.

How Elastic Load Balancing Works:

- Load balancers accept incoming traffic from clients and route requests to registered targets (e.g., EC2 instances) in one or more Availability Zones.
- **Listeners** are configured to check for connection requests on specific protocols and port numbers.
- Load balancers perform health checks to ensure traffic is only routed to healthy targets. If a target becomes unhealthy, it stops receiving traffic until it is healthy again.

Use Cases for Elastic Load Balancing:

- **High Availability and Fault Tolerance:** Ensures that traffic is balanced across healthy targets in different Availability Zones.

- **Containerized Applications:** Supports load balancing across multiple ports on the same EC2 instance, with integration into Amazon ECS for managing Docker containers.
- **Elasticity and Scalability:** Works with Amazon CloudWatch and EC2 Auto Scaling to automatically scale applications based on traffic demands.
- **Virtual Private Cloud (VPC):** Acts as a public entry point into your VPC or routes traffic between application tiers within the VPC.
- **Hybrid Environments:** Allows load balancing across AWS and on-premises resources.
- **Invoking Lambda Functions:** Supports routing HTTP(S) requests to Lambda functions, enabling serverless applications.

Monitoring Load Balancers:

- **Amazon CloudWatch Metrics:** Provides data points to monitor performance and create alarms for specific metrics.
- **Access Logs:** Capture detailed request information for traffic analysis and troubleshooting.
- **AWS CloudTrail Logs:** Capture detailed information about API interactions, including the source of calls and their timestamps.

Key Takeaways:

- Elastic Load Balancing helps distribute traffic and improve application availability.
- The three types of load balancers (ALB, NLB, CLB) serve different needs based on the application architecture.
- ELB includes features for health checks, security, and monitoring, facilitating the creation of resilient applications on AWS.

Section 2 Cloudwatch

Amazon CloudWatch Overview

- **Purpose:** Monitors AWS resources and applications in real-time to provide insight into resource utilization, performance, and operational health.
- **Use Cases:**
 - Determine when to launch additional Amazon EC2 instances.
 - Assess if application performance or availability is affected by capacity issues.
 - Analyze the actual usage of infrastructure.

Key Features of Amazon CloudWatch

- **Monitoring Capabilities:**
 - Monitors AWS resources and applications running on AWS.
 - Collects standard and custom metrics.

- **Alarms:**
 - Can set alarms based on static thresholds, anomaly detection, or metric math expressions.
 - Alarms can trigger notifications to Amazon SNS, initiate Amazon EC2 Auto Scaling actions, or other EC2 actions.
- **Events:**
 - Allows the definition of rules to match operational changes and route events to targets for processing (e.g., EC2 instances, AWS Lambda, Kinesis streams).

Creating CloudWatch Alarms

- **Configuration Options:**
 - Namespace (e.g., AWS/EC2), Metric (e.g., CPU Utilization), Statistic (e.g., average), Period, and Conditions (e.g., Greater than, Less than).
 - Actions can include notifications or Auto Scaling.

Key Takeaways

- Amazon CloudWatch provides real-time monitoring and observability for AWS resources and applications.
- It helps in collecting metrics, setting alarms, and responding to operational changes effectively.

This summary encapsulates the essential aspects of Amazon CloudWatch from the provided content, focusing on its monitoring capabilities, alarm creation, and overall utility in AWS resource management.

Section 3 Auto Scaling:

1. **Provisioned Capacity:**
 - Automatic scaling is beneficial for predictable workloads, such as the weekly traffic patterns seen at Amazon.com.
 - In November, traffic surges significantly due to events like Black Friday and Cyber Monday. Fixed capacity may lead to underutilization, with 76% of resources being idle, thus necessitating dynamic scaling.
2. **Auto Scaling Groups:**
 - An Auto Scaling group is a collection of Amazon EC2 instances that can be scaled automatically.
 - Users can specify minimum and maximum sizes and desired capacity for the group.
 - Scaling policies determine when to launch or terminate instances based on demand.
3. **Scaling Actions:**

- **Scaling Out:** Refers to launching additional instances.
- **Scaling In:** Refers to terminating instances when demand decreases.
- 4. **Dynamic and Predictive Scaling:**
 - **Dynamic Scaling:** Automatically adjusts capacity based on real-time performance metrics (e.g., CPU utilization). This involves using CloudWatch alarms to trigger scaling events.
 - **Predictive Scaling:** Uses historical data and machine learning to forecast demand and scale resources accordingly.
- 5. **Scaling Mechanisms:**
 - **Manual Scaling:** Adjusting capacity by specifying changes in minimum, maximum, or desired instances.
 - **Scheduled Scaling:** Automatically adjusts capacity at defined times based on predictable traffic patterns.
 - **Health Checks:** Periodic checks ensure the desired number of running instances is maintained.
- 6. **Integration with Other AWS Services:**
 - Amazon EC2 Auto Scaling works with Elastic Load Balancing to manage traffic distribution and ensure application availability.
 - AWS Auto Scaling monitors applications across various resources (like ECS tasks and DynamoDB tables) and automatically adjusts capacity to maintain performance at minimal costs.

Takeaways:

- Scaling is essential for responding to changes in resource needs promptly.
- Auto Scaling maintains application availability by dynamically managing EC2 instances.
- Understanding the relationship between Auto Scaling groups, launch configurations, and scaling actions is crucial for effective resource management in AWS.

This summary encapsulates the main themes of Auto Scaling as outlined in the training materials, emphasizing its importance for managing compute resources efficiently.

Module Check

1. Which of the following AWS tools help your application scale up or down based on demand? (Choose two.)

- ☐ Availability Zones
- ☒ Amazon EC2 Auto Scaling
- ☐ AWS CloudFormation
- ☒ Elastic Load Balancing
- ☐ AWS Config

2. Which service would you use to send alerts based on Amazon CloudWatch alarms? (Select the best answer.)

- ☒ Amazon Simple Notification Service (Amazon SNS)
- ☐ AWS CloudTrail
- ☐ AWS Trusted Advisor
- ☐ Amazon Route 53

3. Which of the following are characteristics of Amazon EC2 Auto Scaling? (Choose three.)

- ☐ Only supports dynamic scaling
- ☒ Responds to changing conditions by adding or terminating instances
- ☐ Delivers push notifications
- ☒ Launches instances from a specified Amazon Machine Image (AMI)
- ☒ Enforces a minimum number of running Amazon EC2 instances

4. Which of the following must be configured on an Elastic Load Balancing load balancer to expect incoming traffic? (Select the best answer.)

- ☐ A port
- ☐ A network interface
- ☒ A listener
- ☐ An instance

5. Which of the following elements are used to create an Amazon EC2 Auto Scaling launch configuration? (Choose three.)

- ☒ Amazon Machine Image (AMI)
- ☐ Load balancer
- ☒ Instance type
- ☐ Virtual private cloud (VPC) and subnets
- ☒ Amazon Elastic Block Store (Amazon EBS) volumes

6. Which of the following services can help you collect important metrics from Amazon Relational Database Service (Amazon RDS) and Amazon Elastic Compute Cloud (Amazon EC2) instances? (Select the best answer.)

- ☐ Amazon CloudFront
- ☐ Amazon CloudSearch
- ☒ Amazon CloudWatch
- ☐ AWS CloudTrail
- ☐ Amazon EC2 Auto Scaling

7. Which of the following are elements of an Auto Scaling group? (Choose three.)

- ☒ Minimum size
- ☐ Health checks
- ☒ Desired capacity
- ☒ Maximum size

8. There is an audit at your company and they need to have a log of all access to AWS resources in the account. Which of the following services can assist in providing these details? (Select the best answer.)

- ☐ Amazon CloudWatch
- ☒ AWS CloudTrail
- ☐ Amazon Elastic Compute Cloud (Amazon EC2)
- ☐ Amazon Simple Notification Service (Amazon SNS)

9. In Elastic Load Balancing, when the load balancer detects an unhealthy target, which of the following are true? (Choose three.)

- ☒ Stops routing traffic to that target
- ☐ Triggers an alarm
- ☒ Resumes routing traffic when it detects that the target is healthy again
- ☐ Resumes routing traffic when manually restarted
- ☒ Routes traffic to a healthy target

10. What are the three types of load balancers that Elastic Load Balancing offers?

- ☐ Internet Load Balancer
- ☒ Application Load Balancer
- ☒ Network Load Balancer
- ☐ Compute Load Balancer
- ☒ Classic Load Balancer
- ☐ Auto Scaling Load Balancer