AWS Certified Cloud Practitioner (CLF-C02)

Study Notes

Introduction

- **Introduction**
  - AWS Certified Cloud Practitioner is considered an entry-level certification, and it's going to focus on your understanding of cloud computing concepts including
    - Benefits and drawbacks of cloud computing
    - Different types of cloud services
    - Financial and strategic implications of using cloud services
  - This certification is designed for professionals who are new to cloud computing and professionals seeking an understanding of cloud computing and all its related concepts
  - You will find that the first portion of this course is going to be focused on the big picture theory and concepts related to cloud computing
    - Cloud Computing Principles
    - Cloud Design Principles
    - Cloud Migrations

    - Global Infrastructure
    - Cloud Economics and Pricing
    - Connecting to AWS
  - AWS Certified Cloud Practitioner certification exam consists of 4 domains or knowledge areas
    - 24% of Cloud Concepts
    - 30% of Security and Compliance
    - 34% of Cloud Technology and Services
    - 12% of Billing, Pricing, and Support
  - AWS Certified Cloud Practitioner has a maximum 65 questions over the course of 90 minutes
    - Within the 65 questions, there will be 15 that AWS calls "unscored" questions
    - These unscored questions are new questions that Amazon is evaluating to see if they should be used as real questions in the future
    - When you are taking the exam, you will not know which of the questions are unscored, so it is important to try your best on all of the questions you see on the exam

- To pass the AWS Certified Cloud Practitioner (CLF-C02) exam, you must score at least 700 points out of 1,000 points that are available on the exam
- In order for you to take the exam, you will have to pay an exam fee by buying an exam voucher
    - How do you sign up and schedule your exam?
        - Pearson VUE
            - You can take it at any Pearson VUE testing center worldwide, at either a local testing center or online
            - You can buy that exam voucher by going to Pearson Vue directly when you're scheduling your exam at **pearsonvue.com**, or going to the voucher store at lpi.org to buy it from their online store
            - Pearson VUE and LPI have now created a capability for you to take your certification exam online from the comfort of your home or office, using the Pearson VUE OnVue testing system
        - AWS
            - Go directly to the AWS certification page at **aws.training/certification**
- 4 tips for success in this course
    - Turn on closed captioning
    - Control the playback speed
    - Join our FB group

            - facebook.com/groups/diontraining
        - Download and print the study guide

- **Exam Tips**
    - There will be no trick questions
        - Always be on the lookout for distractors or red herrings
        - At least one of the four listed possible answer choices that are written to try and distract you from the correct answer
    - Pay close attention to words in bold, italics, or all uppercase
    - Answer the questions based on AWS Certified Cloud Practitioner knowledge
        - In the world of project management, there is often not a 100% correct answer to every question you face in your daily work
        - When in doubt, choose the answer that is correct for the highest number of situations
    - Understand the key concepts of the test questions
    - Do not memorize the terms word for word, try to understand them instead
    - During the exam, the answers will be from multiple-choice style questions

- **100% Pass Guarantee**
    - All the risk is on us, as it should be

    - You have nothing to lose here, but you do have to do your part and put in some effort
  - When you take those quizzes, you have to score at least an 80% for it to be considered a pass in our system
  - At the end of the course, you will find our practice exams
    - Understand why the answers are right or wrong
    - Explanations are provided for every single question
  - Please don't try to simply memorize the questions, but instead take the time to understand the why behind them
  - Make sure that you watched the videos, took the quizzes, did the labs, and finished the practice exams
    - If you've done all and don't see the progress part at the top going from 0 to 100, that means something's wrong
    - If you think you've done everything and it still doesn't show 100%, please email us at **support@diontraining.com**
  - Once you have the course completion letter, you are eligible for our 60-Day 100% Pass Guarantee

Cloud Computing Principles

Objectives:

- 1.1 - Define the benefits of the AWS Cloud.
- 1.3 - Understand the benefits of and strategies for migration to the AWS Cloud.
- 2.1 - Understand the AWS shared responsibility model

- **The AWS Cloud**
  - Amazon Web Services is a comprehensive and widely adopted cloud platform that offers over 200 full-featured services from data centers located around the globe
  - AWS provides a vast array of services, ranging from foundational infrastructure technologies
    - Compute
    - Storage
    - Databases
    - Machine Learning
    - Artificial Intelligence
    - Data Lakes
    - Analytics
    - Internet of Things (IoT)

- One of the strengths of AWS lies in the depth of functionality within its services
- *AWS Partner Network (APN)*
  - System integrators and independent software vendors who specialize in AWS services and enables customers to access a diverse ecosystem of expertise and solutions
- AWS is designed to provide flexibility and security
- AWS offers an extensive set of over 300 security, compliance, and governance services and features
- *AWS Regions*
  - Each region is designed to be isolated from the others, ensuring that they are insulated from failures in other region
- *AWS Availability Zones (AZs)*
  - Distinct data centers equipped with their power, cooling, and networking to ensure fault tolerance
- AWS has cemented its place as the go-to platform for businesses seeking to harness the cloud's true potential

- **Benefits of Cloud Computing**
  - *Cloud Computing*
    - Technology that allows for the delivery of computing resources, such as servers, storage, and software, over the Internet

- According to the National institute of Standards and Technology (or NIST) in their Special Publication 800-145 entitled “The NIST Definition of Cloud Computing”, there are five key benefits or features of cloud computing
  - *On-demand Self-service*
    - Key characteristic of cloud computing that enables users to quickly and easily provision computing resources as needed, without the need for human interaction
    - On-demand self-service provides more agility and flexibility by giving the ability to quickly scale up or down
    - Also provides the ability to control the resources by allowing users to easily provision additional resources
  - *Broad Network Access*
    - Allows users to access cloud computing resources from anywhere with an Internet connection
    - The increased mobility also allows organizations to change their business models from a purely in-person model to a hybrid or remote work capable organization
    - Data can be accessed from anywhere which means the businesses can maintain operations at all times
  - *Resource Pooling*
    - Enables cloud providers to dynamically allocate and reallocate resources, such as servers and storage, as needed
    - This leads to increased cost savings, more efficiency, improved scalability, better disaster recovery, and additional flexibility for the operations

      - This resource pooling also allows for better performance and faster service delivery
    - *Rapid Elasticity*
      - Enables resources to quickly and easily scale up or down in response to changes in demand
      - In traditional IT environments, businesses often have to plan ahead and make large upfront investments in infrastructure
      - This can be especially beneficial for businesses that experience seasonal fluctuations in demand
    - *Measured Service*
      - Allows cloud providers to track and bill users for the resources they consume, based on usage
      - Measured service in cloud computing provides businesses with the ability to better track, manage, and optimize their resource usage
      - Businesses can optimize their resource usage patterns to ensure efficient and cost-effective utilization

- **Cloud Service Models**
  - Cloud Service Models
    - Determine the level of control, flexibility, and management required from the user's side
  - 4 Cloud Service Models
    - *Software as a Service (SaaS)*

    - The service provider is going to give the organization a complete solution, including the hardware, as well as the operating system, the middleware, the runtime, the data processing, and the application or software that's needed for the service to be delivered to your end user
- *Infrastructure as a Service (IaaS)*
    - The service provider is going to give the organization the hardware that is needed, including networking, storage, and servers, but they're not going to provide any of the software
- *Platform as a Service (PaaS)*
    - The service provider will provide only the hardware, including the networking, storage, servers and virtualization, as well as the operating system, middleware, and runtime applications
- *Function as a Service (FaaS)*
    - The service provider handles all of the infrastructure, including servers, networking, and storage
    - With FaaS, there is no need to set up or manage servers, configure networks, or handle any of the other infrastructure concerns
    - You get to focus entirely on your application logic, while the service provider handles everything else
    - A great example of a function as a service solution is the AWS Lambda service
        - **AWS Lambda** create functions in a variety of programming languages, deploy them, and then set up triggers such as API Gateway endpoints, S3 bucket events, or DynamoDB stream events

- **Service Model Selection**
    - 3 Main Service Models
        - *Infrastructure as a Service (IaaS)*
            - Provides organizations with access to virtualized computing resources, such as servers, storage, and networking
            - IaaS is a cost-effective way for organizations to access computing resources, and it offers scalability and flexibility
        - *Platform as a Service (PaaS)*
            - Builds on top of IaaS and provides organizations with a platform for developing, testing, and deploying applications
            - PaaS is a cost-effective way for organizations to develop and deploy applications, and it offers scalability and flexibility
        - *Software as a Service (SaaS)*
            - The most advanced of the three service models and provides organizations with access to software applications that can be accessed over the Internet
            - SaaS is a cost-effective way for organizations to access software applications, and it offers scalability and flexibility
    - When selecting a service model, organizations should consider their specific needs, such as scalability, security, and cost
    - Service model selection is not a one-time decision and it can be changed as per the organization's needs

- **Deployment Models**

- Each model has its own set of advantages and disadvantages, and the right model for an organization will depend on its specific needs and requirements
- 6 Types of Cloud Deployment Models
    - *Public*
        - A third-party service provider makes resources available to the end users over the Internet
        - There are numerous public cloud solutions available today including those from Amazon Web Services, Microsoft Azure, and the Google Cloud Platform
        - Public clouds can often be an inexpensive way for an organization to gain a required service quickly and efficiently
        - Advantages of public clouds include
            - Lower costs
            - Ability to scale resources quickly and easily
        - Public clouds can be less secure than private clouds
    - *Private*
        - Cloud computing environment that is dedicated to a single organization that the infrastructure, resources, and data are all exclusively used by that organization
        - The organization is responsible for the design, implementation, and operation of the cloud resources and the servers that host them
        - AWS also offers Virtual Private Cloud (VPC) for users who want the scalability of the cloud while maintaining a private, isolated environment
        - The advantages of private clouds include

        - Increased security and compliance
        - Greater control over the infrastructure and resources
- *Hybrid*
    - Combines the benefits of both the public cloud and the private cloud options
    - Strict rules should be applied for what type of data is hosted in each portion of the hybrid cloud
    - Hybrid clouds have several advantages
        - Increased security and compliance
        - Scale resources quickly and easily
    - Hybrid clouds can be complex to set up and maintain, and organizations may need to invest in additional resources such as hardware and staff
    - This model offers the flexibility of the public cloud with the control and security of a private cloud
- *Community*
    - Cloud computing environment that is shared by a group of organizations with similar requirements
    - The resources and costs are shared among several different organizations who have common service needs
    - AWS GovCloud is a specialized region designed to host sensitive data and regulated workloads in the cloud, meeting U.S. government compliance requirements
    - Community clouds have several advantages
        - Increased security and compliance
        - Ability to share resources and costs among the community

            - The security challenge here is that each organization may have their own security controls
        - *Multi-tenancy*
            - Under this model, the same resources are used by multiple organizations
            - When two or more organizations share the same resources, there are some security concerns
        - *Single tenancy*
            - Under this model, a single organization is assigned to a particular resource
            - Single tendency is much less efficient than multi-tendency, and it is also more expensive because it requires more hardware to run properly

- **Deployment Model Selection**
    - The choice of model significantly influences the cost, control, scalability, and security of applications and services
    - *On-premise Deployment*
        - Involves hosting applications and data on the infrastructure located within the organization's premises
        - It requires significant upfront capital investment and ongoing maintenance costs
        - Private cloud deployment model is usually paired with an on-premise deployment solution

- *Cloud Deployment*
  - Involves hosting applications and data on a cloud platform
  - Public cloud deployment is also referred to as a cloud deployment
- *Hybrid Deployment*
  - Combines on-premises and cloud-based infrastructures into one deployment model
  - If an organization has substantial on-premise infrastructure but wants to leverage the benefits of the cloud, a hybrid model could be a good fit
- *Multi-cloud Deployment*
  - Involves using more than one cloud service from different providers
  - Organizations that are seeking to avoid vendor lock-in or to meet specific geographic or service requirements may opt for a multi-cloud approach
- Remember, the choice of deployment model depends on various factors
  - Business needs
  - Budget
  - Compliance requirements
  - Technical capabilities
- The key to successful deployment model selection is aligning the model with your organization's strategic objectives and operational needs

- **Shared Responsibility Model**
    - *Shared Responsibility Model*
        - Framework used in cloud computing to define the roles and responsibilities of the cloud service provider and the customer when it comes to security and compliance in the cloud
    - Basic Principle of Shared Responsibility Model
        - Cloud Service Provider
            - Responsible for the security of the cloud itself
        - Customer
            - Responsible for the security of their own data and applications
    - *Amazon Web Services (AWS)*
        - Cloud service provider that provides a variety of cloud services, including storage, computing power, and databases
        - Customer is responsible for securing their own data and applications
        - AWS is responsible for the security of the underlying infrastructure including:
            - Network security
            - Physical security of the data centers
            - Security of the virtualization layer

        - As Amazon Web Services puts it, the customer is responsible for security "in" the cloud, while AWS is responsible for the security "of" the cloud
        - AWS must manage the security of the cloud, ensuring the robustness and reliability of the provided infrastructure
    - Customers are responsible for the security of their usage of the AWS Cloud services
        - Customers need to handle aspects like
            - Client and endpoint protection
            - Encryption of data at rest and in transit
            - Network traffic protection
            - Managing access control to AWS resources
        - For services that are more abstracted, such as Amazon's S3 and DynamoDB, AWS manages the infrastructure layer, the operating system, and the platform
    - There are also some shared responsibilities between AWS and the customer
    - The division of responsibilities can shift depending on the AWS service used
        - Customers retain responsibility for the security of their applications and data
    - One advantage
        - The shared responsibility model can also help to reduce the overall cost of security
    - Some disadvantages of the shared responsibility model

        - It can be difficult for customers to fully understand their responsibilities and to properly secure their own data and applications
        - It can create confusion or ambiguity about who is responsible for specific security tasks
    - The shared responsibility model can be an effective way of ensuring that both the cloud service provider and the customer are taking the necessary steps to secure their respective parts of the cloud environment

Cloud Design Principles

Objectives:

- 1.1 - Define the benefits of the AWS Cloud.
- 1.2 - Identify design principles of AWS Cloud.

- **Scaling**
  - *Scaling*
    - Process of increasing or decreasing the resources of a cloud system to meet the changing needs of an organization
  - There are two common types of scaling methods used in cloud designs
    - *Vertical Scaling (Scaling Up or Down)*
      - Involves increasing or decreasing the resources of a single server
    - *Horizontal Scaling (Scaling Out or In)*
      - Involves adding or removing servers from a system to increase or decrease resources
      - Horizontal scaling is flexible to increase or decrease resources, but it is more complex and time-consuming to set up and manage
  - When selecting a scaling strategy, organizations should consider their specific needs
    - Scalability
    - Cost
    - Ease of Management

- Differences between Vertical and Horizontal Scaling

Vertical Scaling	Horizontal Scaling
Easy to manage	More scalable
Cost-effective	Can handle more traffic
Increase an individual node's performance	Can add more nodes as demand increases
Efficient	Easy to balance between nodes' workload

- Drawbacks to horizontal scaling
  - Horizontal scaling can be more complex to manage when it comes to configuring and maintaining additional nodes
  - Horizontal scaling can be more expensive when adding additional nodes to the system
- When choosing between horizontal and vertical scaling, there are several factors to consider
  - Type of workload
  - Expected growth rate
  - Cost

- **Redundancy**

- *Redundancy*
    - The process of creating multiple copies of data or resources to ensure the availability and reliability of systems
    - Redundancy helps to ensure the continuity of operations, reduce downtime, and prevent data loss
- There are several types of redundancies that are commonly used in cloud designs
    - *Data Redundancy*
        - Duplication of data in multiple locations, such as storage devices or servers
    - *Hardware Redundancy*
        - Use of multiple physical components, such as servers or storage devices
    - *Network Redundancy*
        - Use of multiple network connections to ensure systems continue to operate during a network failure
- Redundancy helps to ensure the availability and reliability of systems
- There are five main benefits to using redundancy in your cloud designs
    - High Availability
        - Helps to ensure that systems continue to operate when failure occurs
    - Fault Tolerance
        - Provides a backup in case of failures or errors, which helps to prevent data loss and minimize downtime

        - Improved Performance
            - Provides a backup system in case of a failure
        - Increased Resiliency
            - Provides a backup or secondary system in case of failures
        - Improved Security
            - Helps to improve the security of systems by providing a backup system in case of failures

- **High Availability**
    - *High Availability*
        - The ability of a system to remain operational and accessible to users with minimal downtime and is considered a measure of the system's reliability
        - This high availability is achieved through the use of redundant systems, network connections, and other components that can take over operations in case of a failure
    - There are several strategies that organizations can implement to achieve high availability in cloud computing
        - *Load Balancing*
            - Distributing incoming traffic across multiple servers to ensure that no server is overwhelmed
        - *Failover*

            - Having secondary systems that can take over in the event of a failure
        - *Disaster Recovery*
            - Having a plan in place to recover from unexpected events
    - There are five key benefits of using high availability in your cloud designs
        - Minimized downtime
        - Improved performance
        - Increased resilience
        - Improved security
        - Enhanced user experience
    - 4 key factors or events that can affect the levels of high availability we can achieve in our cloud designs
        - Network Connections
            - Critical component of cloud computing where failures can result in significant downtime
        - Server and Storage Failures
            - Can result in significant downtime and data loss
        - Power Outages
            - Ensure backup power sources are available to provide redundancy
        - Natural Disasters

            - Ensure that systems are designed to withstand natural disasters and that backup systems are available in case of a disaster
    - When implementing high availability in cloud computing, organizations should consider their specific needs, such as availability, cost, and ease of management

- **Disaster Recovery**
    - *Disaster Recovery (DR)*
        - The process of restoring IT systems, applications, and data after a disruption
    - There are several strategies that organizations can implement to achieve disaster recovery in cloud computing
        - Backup and Restore
            - Creates regular backups of data and applications and restores them in the event of a disaster
        - Replication
            - Creates a second copy of the system and takes over in the event of a disaster
        - Failover
            - Having a secondary system that can take over in the event of a disaster
    - Disaster recovery ensures the system is available and accessible during a disaster
    - Cloud computing provides several advantages for disaster recovery

        - Scalability
            - Organizations can increase or decrease computing resources as needed
        - Flexibility
            - Organizations can respond to changes in disaster recovery needs
        - Cost Savings
            - Organizations can pay only for the resources they need
        - Improved Data Protection
            - Organizations can recover from data loss or corruption more quickly
    - 3 main types of disaster recovery strategies used in cloud computing
        - *Hot Site*
            - Fully-functional alternate location that is constantly maintained with the latest hardware, software, and data
            - A hot site is best suited for organizations with critical operations that require quick and seamless recovery
            - Provides real-time data replication and immediate access to critical applications and data during a disaster
        - *Warm Site*
            - Partially-functional alternate location with some infrastructure and equipment in place, but doesn't have all the latest hardware and software

        - Warm sites offer a lower-cost alternative and can provide recovery time, but require more preparation and planning
        - Provides partial infrastructure and data, which allows for recovery from a disaster in a shorter amount of time
    - *Cold Site*
        - Pre-prepared location with basic infrastructure, such as power and network connections, but without any hardware, software, or data
        - Cold sites offer the lowest cost option, but require the most preparation and planning and have the longest recovery times
        - Stores backups and snapshots of critical data and allows for data recovery during a disaster
  - When implementing disaster recovery in cloud computing, organizations should consider their specific needs
    - Availability
    - Cost Ease of Management
    - Recovery Time Objectives
    - Overall Disaster Recovery Strategy
  - Review and evaluate the disaster recovery plan to ensure it continues to meet the organization's needs

- **Recovery Objectives**
  - *Recovery Objectives*

        - Desired state of an organization's systems, applications, and data after a disaster or other disruption
    - Recovery objectives are critical components of disaster recovery plans and play a crucial role in cloud-based designs
    - 3 Main Recovery Objectives
        - *Recovery Time Objective (RTO)*
            - Time it takes to recover critical systems and data after a disaster occurs
        - *Recovery Point Objective (RPO)*
            - Maximum amount of data that can be lost during a disaster
        - *Maximum Tolerable Downtime (MTD)*
            - Maximum amount of time that critical systems and data can be unavailable without causing significant harm to the organization
    - Recovery objectives can be achieved through real-time data replication, backup and restore, and failover
        - If you have a low RTO and RPO
            - Use a hot site
        - If you have a high MTD but low RPO
            - Use a warm site
    - Reviewing recovery objectives regularly ensures organizations remain in alignment with their needs and risk tolerance

- **AWS Well-Architected Framework**

- 6 basic principles (Key Pillars) that lead to building successful cloud infrastructure and applications
  - *Operational Excellence*
    - Focuses on managing and automating changes, monitoring systems, and continuously improving procedures
    - Application designs should also include infrastructure as code, performance monitoring, and incident response planning considerations
    - Using something like CloudWatch, we can set up custom metrics and alarms to proactively identify performance bottlenecks, resource constraints, and anomalies
  - *Security*
    - Focuses on the protection of data, infrastructure, and assets
  - *Reliability*
    - Systems will maintain stability, recover from failures, and consistently meet customers' needs
    - AWS CloudWatch enables monitoring of AWS resources, applications, and custom metrics
  - *Performance Efficiency*
    - Involves allocating the right resources to meet application requirements, maintaining responsiveness, and optimizing costs
    - AWS offers services like Amazon CloudFront, Amazon ElastiCache, and Amazon S3 to cache and deliver content efficiently
  - *Cost Optimization*
    - Ensures that organizations obtain maximum value from their resources while minimizing costs

            - AWS provides tools like AWS Cost Explorer and AWS Budgets to monitor and analyze spending patterns
        - *Sustainability*
            - Addresses the increasing importance of making business better for the environment
            - AWS Lambda and other serverless services eliminate the need for provisioning and managing servers
    - To help companies address the six pillars, AWS offers the Well-Architected Tool
        - *Well-Architected Tool*
            - Comprehensive and invaluable resource for assessing and optimizing cloud architectures to align with AWS best practices and against the six pillars to identify ways to improve the architecture
            - This tool offers a series of questions and guidelines that help users review their cloud infrastructure and identify potential areas for improvement

- **AWS Well-Architected Tool**
    - *AWS Well-Architected Tool*
        - Service in the AWS Management Console that helps review the workload against AWS architectural best practices
    - The tool will ask you questions from across each of its six pillars
        - Operational Excellence
        - Security

- Reliability
- Performance Efficiency
- Cost Optimization
- Sustainability

Cloud Migrations

Objective 1.3 - Understand the benefits of and strategies for migration to the AWS Cloud.

- **Migrating to the Cloud**
  - *Cloud Migration*
    - The process of moving data, applications, and other business processes from one environment to another
  - *On-premise environment to a cloud-based environment*
    - On-premise environments are data centers and servers that are physically located within a company's own premises, while cloud-based environments are servers that are maintained and owned by third-party cloud service providers who make their services available to their customers over the Internet
    - Advantages
      - Improved scalability
      - Increased accessibility
      - Reduced costs
    - Disadvantages
      - Concerns surrounding the security of the organization’s data
      - The loss of control over IT infrastructure and data can be a significant concern for many organizations
  - *Cloud-based environment to cloud-based environment*

- Data and applications are transferred from one cloud service provider to another
- Advantages
  - Cost savings
  - Migration can lead to improved performance and reliability
  - Ability to take advantage of new features and services offered by the new provider
- Disadvantages
  - Time and effort required to move all the data and applications to the new cloud environment
  - There may be compatibility issues between the old and new cloud environments
  - Many cloud providers have large data transfer fees when moving from one cloud provider to another

- *Cloud environment to on-premise environment*
  - A cloud-to-on-premise migration process involves transferring data and applications from a cloud service provider to an in-house data center, typically for reasons such as increased security, control, and customization
  - Advantages
    - Companies have complete control over their security measures and can tailor them to meet their specific needs

- Increased level of control over the IT infrastructure
- Disadvantages
  - Large upfront cost of setting up and maintaining an on-premise infrastructure
  - The ongoing maintenance and management of the infrastructure can also be time-consuming and expensive

- **Cloud Migration Phases**
  - *Cloud Migration*
    - Process of moving data, applications, or other business elements from an on-premises environment to the cloud
  - Process
    - Planning phase sets the foundation for a successful migration into the cloud
      - Conducting an Initial Assessment
        - The organization will assess its current IT infrastructure and identifies the applications, data, and resources that need to be migrated to the cloud
      - Conducting a Cloud Readiness Assessment
        - The organization will assess its readiness for cloud migration, including its technical, financial, and organizational readiness
      - Creating a Business Case
        - It will be responsible for outlining the reasons for migrating to the cloud and the benefits
      - Developing a Cloud Strategy

        - Outline the steps it will take to migrate to the cloud, including the timeline, resources, and budget required
    - Selecting a Cloud Provider
        - Determine which cloud provider can best meet your organization's needs while staying within any budgetary constraints may be facing

- The implementation phase involves the actual migration of data, applications, and other resources to the cloud
    - Data Migration
        - Organization will migrate its data to the cloud, including its databases, files, and other data sources
    - Application Migration
        - Organization should migrate its applications to the cloud, including its customer relationship management systems, productivity suites, accounting systems, and other critical applications used by the organization on a daily basis
    - Infrastructure Migration
        - Organization will migrate its infrastructure to the cloud, including its servers, storage, and networking components
    - Testing
        - Organization will test the migrated data, applications, and infrastructure to ensure that everything is working as expected and this will finish up the implementation phase

- Optimization and security phase involves optimizing the migrated cloud environment and ensuring that it is secure from attack
    - Performance Optimization
        - Organization should be focused on optimizing the performance of its cloud environment, including the applications, data, and infrastructure
    - Cost Optimization
        - Company is focused on reducing unnecessary expenses and optimizing its usage of cloud resources
    - Improving the security
        - Organization will implement security measures to protect its cloud environment, including firewalls, intrusion detection and prevention systems, and encryption

- **Cloud Migration Types**
    - 8 Different Types of Cloud Migrations
    - *Rehost (Lift and Shift)*
        - Simplest and quickest type of cloud migration involves moving an existing app from an on-premises environment to a new cloud environment without changes to the application's code or architecture
        - Several advantages to a rehosting migration
            - Quick and simple

- Minimal downtime
- Cost-effective

- Few disadvantages to conducting a rehosting migration
  - Rehosting provides limited benefits
  - Rehosting can result in an inflexible infrastructure that is not optimized for the cloud

- *Replatform* (Lift, Tinker, and Shift/Re-architecture Migration)
  - Involves making changes to the architecture of an existing application
  - Several advantages to a replatform migration
    - Improved performance
    - Cost savings
    - Improved flexibility
  - Some downsides to using replatforming
    - Replatforming can be time-consuming and requires expertise
    - Replatforming can be a high-cost option for organizations
- *Refactor* (Rip and Replace)
  - Involves breaking down an existing application into smaller and more manageable components (microservices)
  - Several advantages to a refactoring migration
    - Improved performance
    - Scalability
    - Cost savings

        - Some disadvantages of using refactoring
            - Refactoring can be complex and requires expertise
            - Refactoring is time-consuming
    - *Repurchase* (Drop and Shop)
        - Involves replacing an existing application with a commercially available Software-as-a-Service (SaaS) solution
        - Several advantages to a repurchase migration
            - Quickness and simplicity
            - Cost-effectiveness
            - Improved security
        - Few disadvantages to using repurchasing
            - Repurchasing can result in limited customization options
            - Repurchasing can result in dependence on the vendor for maintenance and support
    - *Retire*
        - Involves discontinuing an existing application and replacing or discontinuing the use of the old product without replacing it with a new one
        - Advantages
            - Improved security
            - Cost savings

- Disadvantages
  - Retiring can result in data loss
  - Retiring can result in disruption to business operations
- *Retain*
  - Involves keeping applications or systems in their present environment because legal, regulatory, or other factors may prevent a migration to the cloud at this time
  - Several advantages to a retain migration
    - No modifications
    - No disruption
  - Couple of disadvantages to using retain
    - Retaining does not provide us with any cost savings
    - We cannot achieve any cloud-based performance, security, or cost improvements since our legacy application will remain in use and we will have to continue to support it in our on-premise environment
- *Hybrid*
  - Involves using a combination of different migration types to move an organization's applications and data to the cloud
  - Several advantages to a hybrid migration
    - Customized solution
    - Cost savings

    - Some disadvantages to a hybrid migration
        - Hybrid migrations can be complex
        - Hybrid migrations can be time-consuming
  - *Phased*
    - Involves moving applications and data to the cloud in stages
    - Several advantages to a phased migration
        - Minimized risks
        - Improved planning
    - Few disadvantages to using a phased migration approach
        - Phased migrations can be time-consuming
        - Phased migrations can increase costs

- **Cloud Native Applications**
  - *Cloud Native Application*
    - Used to build highly scalable, resilient, and flexible apps that can be deployed on cloud infrastructure
    - Cloud native applications make it possible to run applications that are highly available
    - Cloud native applications are designed to be agile and flexible
  - Cloud native applications are usually built using containers and microservices
    - *Container*

            - Lightweight form of virtualization that allows applications to run in an isolated environment
            - Containers provide a consistent runtime environment
        - *Microservice*
            - Way of breaking down a single large program or application into smaller, independent services that can be deployed and managed separately
    - Advantages and Disadvantages
        - Several advantages to using cloud native applications
            - *Scalability*
                - Increase capacity as needed
            - *Resilience*
                - Automatic failover and self-healing features
            - *Flexibility*
                - Easy to change and provides ability to deploy quick updates
            - Cost Savings
                - Deploys efficiently and with less overhead by using containers and microservices
        - Several disadvantages to using cloud native applications
            - Complexity
            - Security issues
            - Skill required

- Organization's cloud native applications should be managed by development and operations teams
  - *DevOps*
    - Culture, mindset, and set of practices that are designed to help organizations to deliver high-quality software faster and more efficiently
    - It is based on the 3 principles
      - Continuous Integration
      - Continuous Delivery
      - Continuous Deployment
  - Cloud native apps require DevOps as they are designed to be scalable, resilient, and flexible

- **AWS Cloud Adoption Framework (CAF)**
  - *AWS Cloud Adoption Framework (CAF)*
    - Comprehensive guide designed by Amazon Web Services to help businesses transition smoothly to cloud-based operations
  - The AWS Cloud Adoption Framework is divided into six different perspectives, or areas of focus, each with its specific role and stakeholders
    - *Business Perspective*
      - Focuses on business cases, benefits realization, and changes to the business model
    - *People Perspective*

        - Involves changes that staff members need to adapt to (New Roles, Organizational Structures, and Skills)
      - *Governance Perspective*
        - Focuses on program and project management, risk management, and business performance measurement
      - *Platform Perspective*
        - Involves the management and provisioning of cloud infrastructure
      - *Security Perspective*
        - Involves data protection, compliance, and identity management
      - *Operations Perspective*
        - Focuses on running workloads (Change Management, Resource Tracking, and Reporting)
    - There are four main benefits of using the Cloud Adoption Framework
      - Reduced Business Risk
        - Provides a structured approach to cloud adoption and reduces the risk of missing crucial steps
      - Improved Environmental, Social, and Governance Performance
        - Governance Perspective promotes transparency and accountability, enabling organizations to meet their environmental, social, and governance performance objectives
      - Increased Revenue
        - Companies can access a global market, operate 24/7, and scale quickly to meet changes in demand

        - Increased Operational Efficiency
            - Promotes efficient operations by providing guidance on the best practices of an organization's cloud operations
    - To utilize the AWS Cloud Adoption Framework, you will progress through four phases
        - Envision Phase
            - Identify and prioritize transformation opportunities according to the organization's own strategic objectives
        - Align
            - Identify any capability gaps and cross-organizational dependencies
        - Launch
            - Implement pilot programs within the live production environment and show proposed changes
        - Scale
            - Broaden the scope of pilot programs and business value to achieve the desired scale for the effort

- **AWS Snow Family**
    - *AWS Snow Family*
        - Suite of physical devices that facilitate data migration to the cloud by transporting large volumes of data into and out of the AWS cloud

- The Snow Family was designed for edge computing and data transfer, and it includes three primary product lines
    - *AWS Snowcone*
        - Smallest member of the Snow Family, and it is designed to be a portable, rugged, and secure edge computing and data transfer device
        - Snowcone supports AWS IoT Greengrass and can run edge computing workloads that use AWS Lambda functions
    - *AWS Snowball*
        - Larger data transfer device, available in two options
            - Snowball Edge Storage Optimized
                - High storage capacity, offers up to 80 TB of hard disk drive or up to 210 TB of SSD
            - Snowball Edge Compute Optimized
                - Provides up to 104 vCPUs, 416 GB of memory, 28 TB of SSD, and an optional GPU for use cases
        - Snowball devices can migrate large volumes of data from on-premises data centers to AWS
        - Snowball device uses encryption for data protection
    - *AWS Snowmobile*
        - Petabyte-scale data transfer service used to move large amounts of data to AWS

AWS Snowcone	**AWS Snowball**	**AWS Snowmobile**

Compact/Portable	Edge Computing/ Edge Storage	Largest device
8 to 14 TB	Configured/Optimized	Petabytes (Data)

- **AWS Migration and Transfer Tools**
    - 7 different migration and transfer tools provided within AWS
    - *AWS Application Discovery Service*
        - Helps organizations plan migration projects by gathering information about on-premises data centers
    - *AWS Application Migration Service*
        - Used to simplify the process of migrating applications from on-premises data centers to the AWS cloud (Rehosting, Re-platforming, and Refactoring)
        - AWS Application Migration Service minimizes cutover windows and reduces the risk of migration failures
    - *AWS Database Migration Service (DMS)*
        - Used to help migrate databases to AWS quickly and securely
        - The AWS Database Migration Service supports
            - homogenous migrations such as an Oracle database to an Oracle database

- heterogeneous migrations between different database types, like MySQL to Amazon Aurora

- *AWS Migration Hub*
  - Provides a single location to track the progress of application migrations across multiple AWS and partner solutions
- *AWS Schema Conversion Tool (AWS SCT)*
  - Makes migrating different database platforms more predictable
  - Any code that cannot be automatically converted is marked to be manually converted by technicians or analysts
- *AWS Snow Family*
  - Consists of the Snowcone, Snowball, and Snowmobile devices, designed to handle data transfer and edge computing tasks in various scenarios
- *AWS Transfer Family*
  - Provides fully managed support for file transfers directly into and out of Amazon S3 or Amazon EFS

- **Performing a Lift and Shift**
  - *Lift and Shift Method*
    - Cloud migration strategy in which an organization moves its existing applications and infrastructure to the cloud without making significant changes to the underlying architecture or design

- It might not be the most efficient or cost-effective method as it does not take advantage of the cloud native features and capabilities

Global Infrastructure

Objective 3.2 - Define the AWS global infrastructure

- **AWS Regions**
  - This infrastructure is organized physically into AWS Regions, and each region is made up of three or more Availability Zones
  - *AWS Regions*
    - Physical location in the world where AWS clusters multiple data centers together
    - AWS has established Regions all around the world to provide a reliable and scalable cloud computing environment
    - These regions allow customers to deploy their applications and data close to their end-users
  - The AWS Cloud contains over 30 geographic regions across the globe, including over 100 availability zones within those regions
    - Each of the different AWS Regions is designed to be completely isolated from the other AWS Regions to ensure the greatest possible fault tolerance and stability
  - When choosing the AWS Region to use, you will generally consider several factors
    - *Latency*

        - The time it takes for data to travel from one point to another, typically measured in milliseconds
        - The shorter the distance between the customers and the AWS Region, the lower the latency their systems will experience
    - Cost
        - The cost and pricing for AWS services vary between different regions
    - Service Availability
        - AWS strives to provide a wide range of services in all regions, but it is also important to note that not all regions support all services
    - Data Sovereignty and Compliance
        - By selecting the appropriate Region, customers can comply with data residency requirements and ensure that their data remains within the jurisdiction where it was created
    - Resiliency
        - For high availability and disaster recovery purposes, deploy the applications across multiple regions
- To select the region you want to use, you can use the following
    - AWS Management Console
    - AWS Command Line Interface
    - AWS Software Development Kits

        - AWS Application Programming Interfaces
    - AWS Regions are used to provide customers with the flexibility to place their resources and applications close to their end users

- **Availability Zones**
    - *Availability Zones (AZs)*
        - Distinct, physically separated locations within a region, each consisting of one or more data centers equipped with independent power, cooling, and networking to ensure fault tolerance
        - These availability zones provide an inexpensive, low-latency network connection to other zones in the same region
        - AWS spans over 100 AZs across 30-plus geographically dispersed regions
    - Choosing an Availability Zone is typically based on
        - Latency
            - Choose Availability Zones that offer the lowest latency to the users
        - Cost
            - Some AWS resources might cost more depending on the Availability Zone
        - Service Availability

            - AWS strives to offer all services in every Availability Zone, but some services might not be available in all Availability Zones within a region
        - Resiliency
            - In order to best design the cloud-based systems for high availability and disaster recovery, distribute the resources across multiple Availability Zones within a given region
    - AWS Availability Zones provide the infrastructure building blocks for creating resilient and fault-tolerant applications

- **Edge Locations**
    - *Edge Location*
        - Physical site or data center located in major cities and highly populated areas across the globe that AWS uses to cache and deliver content
        - AWS Edge Locations are part of the Amazon CloudFront Content Delivery Network (CDN) service, which is designed to distribute content globally with low latency and high data transfer speeds
    - *Amazon CloudFront*
        - Fast content delivery network service that securely delivers data, videos, applications, and APIs to customers globally with low latency and high transfer speeds

        - CloudFront uses Edge Locations to cache copies of the content closer to the users
    - *AWS Global Accelerator*
        - Networking service that sends the user's traffic through Amazon Web Service's global network infrastructure to improve the application's availability and performance
        - When the service is enabled, the user traffic enters AWS's network at the Edge Location that is geographically nearest to the user
        - Caching mechanism helps reduce the load on the origin server, decrease latency, and increase the overall performance of web applications
    - There are several benefits to using AWS Edge locations
        - Decreased Latency and Increased Performance
            - By reducing latency by serving cached content from the Edge Locations closest to the users, AWS significantly reduces latency and provided a smoother, more responsive user experience
        - Increased Scalability
            - AWS edge locations handle traffic spikes and distribute the load across multiple locations
        - Decreased Costs

        - Using Edge Locations, CloudFront, and the Global Accelerator is cheaper than serving up all the content directly from the origin servers
    - Enhanced Security
        - AWS Edge Locations offer Distributed Denial of Service (DDoS) protection, secure data transfer using HTTPS, and the ability to integrate with the AWS Web Application Firewall and the AWS Shield
    - Increased Global Reach
        - Edge Locations are spread across the globe, users from anywhere can access content quickly and efficiently with consistent and reliable performance
    - Customization
        - With Edge Locations, users can configure caching behaviors, create custom error pages, and implement access controls based on geographic locations
  - AWS Edge Locations play a vital role in improving the performance, scalability, and security of web applications

- **AWS Local Zones**
  - *AWS Local Zones*
    - Designed to bring AWS services closer to specific geographical areas

    - The primary goal is to cater to applications that require single-digit millisecond latencies by end-users or on-premises facilities
  - Each local zone is considered to be an extension of its parent AWS Region, but these local zones are located in different geographical areas
  - To use an AWS Local Zone
    - Set up the local zone by enabling them in the Virtual Private Cloud settings from within the AWS Management Console
    - Deploy the applications and resources into the Local Zone
    - Connect the Local Zone to the parent AWS Region through a high-bandwidth and low-latency network link
    - Conduct Resource Management
  - There are many benefits to using Local Zones
    - Reduced Latency
      - By situating AWS services closer to the end-users and on-premise installations, data requests are processed and delivered faster, significantly improving user experience for latency-sensitive applications
    - Seamless Integration
      - Local Zones are seamlessly integrated with AWS Regions, allowing to manage resources, transfer data, and leverage AWS services with ease

        - Cost Savings
            - Local Zones can contribute to cost savings by minimizing data transfer over long distances
        - Improved Scalability
            - Local Zones allow to scale applications efficiently by distributing the load across multiple locations, resulting in having high performance even during traffic spikes
        - Flexible Deployment
            - Local Zones offer the flexibility to run entire applications or specific parts of applications closer to end-users, depending on the performance requirements
        - Compliance and Data Residency
            - By deploying applications in Local Zones, you can meet data residency and compliance requirements by storing and processing data within specific geographic areas
    - AWS Local Zones provide an effective solution to enhance application performance by deploying resources closer to the end-users

- **AWS Wavelength Zones**
    - *Wavelength Zones*

        - Infrastructure deployments that embed AWS compute and storage services within the telecommunications service providers' datacenters at the edge of the 5G network
    - This dramatically reduces the latency for use cases like
        - Machine Learning
        - Industrial Equipment
        - Smart Cars
        - Smart Cities
        - Internet of Things (IoT)
        - Virtual Reality
    - To use AWS Wavelength Zones
        - Set them up by creating resources in the Wavelength Zone just like they would in any other AWS Region
        - The developers must deploy applications and resources in a Wavelength Zone to provide low-latency access to users
        - The Wavelength Zones are connected to the parent AWS Region through a high-bandwidth and low-latency network link

- The developers can manage resources in a Wavelength Zone using the same APIs and AWS Management Console they use for AWS Regions

- Once configured, you will find many benefits to using these AWS Wavelength Zones
  - Ultra-low Latency
    - Wavelength Zones enable ultra-low latency applications by extending AWS infrastructure to the edge of the telecommunications network
  - Seamless Integration
    - Wavelength Zones are seamlessly integrated with AWS Regions and this allows developers to manage resources, transfer data, and use AWS services with ease
  - Improved Application Performance
    - By reducing latency, Wavelength Zones help improve the performance of applications such as online gaming, augmented and virtual reality, and video streaming to provide a better user experience
  - Cost Savings
    - By offloading traffic to Wavelength Zones instead of getting the data from the origin servers each time, it can reduce the organization's data transfer costs
  - Innovation at the Edge

        - Wavelength Zones enable new types of applications that require ultra-low latency, such as real-time analytics, autonomous vehicles, and IoT applications
    - Flexible Deployment
        - The organization can have more optimized performance and cost-effectiveness over using a region-based server on its own
  - AWS Wavelength Zones provide a powerful solution for developers to build and deploy ultra-low latency applications at the edge of the 5G network

- **AWS Outposts**
  - AWS Outposts
    - Fully managed service that extends AWS infrastructure, services, APIs, and tools to virtually any data center, co-location space, or on-premises facility for a truly consistent hybrid experience
    - Outposts allow customers to run compute and storage on-premise, while seamlessly connecting to AWS's broad array of services in the cloud
  - To use AWS Outposts
    - Have AWS install the Outposts hardware and software infrastructure in your on-premise environment
    - AWS Outposts connect to the nearest AWS Region over a secure, high-speed, low-latency network connection

    - Deploy applications and workloads on your Outpost just as you would in an AWS Region
- There are numerous benefits to using AWS Outposts
    - Hybrid Experience
        - AWS Outposts are used to bring native AWS services, infrastructure, and operating models to virtually any data center, co-location space, or on-premise facility
    - Reduced Latency
        - Drastically reduced latency when using AWS Outposts over a more traditional cloud-based deployment model
    - Data Residence Requirements
        - Sensitive data can be processed on-premise within AWS Outposts and stored in whichever AWS Region, in order to meet data residency requirements
    - Integration with AWS services
        - Enable advanced capabilities like data analytics, machine learning, and the usage of the Internet of things
    - Fully Managed Solution
        - AWS takes care of all the maintenance, updates, and infrastructure management for Outposts
- AWS Outposts provide a flexible, fully managed hybrid cloud solution for businesses that need to run applications on-premise and in the AWS cloud

- **Maintaining Reliability, Redundancy, and Availability**
    - In the world of cloud computing, three key terms are often used to measure the quality and performance of cloud-based services
        - *Reliability*
            - Capacity of a system to recover from infrastructure or service disruptions
        - *Redundancy*
            - Involves duplicating key components or functions within a system to increase its reliability
        - *Availability*
            - Refers to the extent to which a system is operational and accessible to users within a given timeframe
    - AWS Auto Scaling
        - Allows applications to dynamically adjust capacity to maintain steady, predictable performance at the lowest possible cost
    - Increasing Redundancy
        - One of the best strategies for ensuring redundancy in AWS is distributing workloads across multiple Availability Zones within a Region
    - Increasing Availability
        - In terms of availability, AWS achieves this by leveraging multiple Availability Zones within a Region for failover redundancy to ensure that your applications remain operational even if one zone goes down

        - Another way to increase availability and performance is to utilize load balancing across multiple resources
            - Elastic Load Balancer
                - Automatically distributes incoming application traffic across multiple targets
    - When working with AWS services, it is important to understand the different AWS infrastructure components such as
        - Regions
            - Geographical areas that consist of multiple, isolated data centers called availability zones
        - Availability Zones (AZs)
            - Individual data centers within a region, providing fault tolerance and high availability
        - Edge Locations
            - Data centers located outside of AWS regions that provide end users with minimal latency and high data transfer speeds
        - Local Zones
            - Extensions of regions that provide infrastructure closer to large population centers to deliver ultra-low latency applications
        - Wavelength Zones

- Infrastructure deployments that extend AWS regions to telecommunication networks that enable ultra-low latency applications for mobile and 5G devices

Cloud Economics and Pricing

Objectives:

- 1.4 - Understand concepts of cloud economics.
- 4.1 - Compare AWS pricing models.
- 4.2 - Understand resources for billing, budget, and cost management.

- **Fixed vs Variable Costs**
    - There are two main categories of expense
        - *Fixed Expenses*
            - Costs that remain constant over a given period, regardless of the level of usage or production
        - *Variable Expenses*
            - Costs that change based on usage or production levels
    - Considerations when examining costs
        - Return on Investment (ROI)
            - Calculated by subtracting the total cost of the project from the total benefits and dividing the result by the total cost of the project
        - Total Cost of Ownership (TCO)
            - Takes into account both fixed and variable costs and ongoing costs associated with maintaining and supporting the project
    - Take advantage of
        - Process Transformation

            - Allows businesses to streamline and automate business processes
        - Business Transformation
            - Allows businesses to create new revenue streams and business models
    - Capital Expenditures (CAPEX) and Operational Expenditures (OPEX)
        - Capital Expenditures (CAPEX)
            - Expenses incurred by a company for the acquisition of long-term assets, (land, buildings, equipment, and technology)
            - CAPEX is used to improve the company's operations and grow its business
        - Operational Expenditures (OPEX)
            - Costs incurred by a company in its day-to-day operations
            - In the context of cloud computing, operating expenditures may include costs associated with using cloud services
                - Storage
                - Computing
                - Data transfer
                - Cloud infrastructure management
                - Employee training
        - In traditional IT infrastructure, investments in hardware and software were Capital Expenditures (CAPEX)

        - In contrast, cloud computing operates on a pay-as-you-go model, where costs are treated as Operating Expenditures (OPEX)
        - The shift from CAPEX to OPEX allows organizations to be more flexible and agile in technology investments
            - Cloud providers handle maintenance and upgrades
    - Remember that fixed and variable expenses are two distinct types of costs that businesses incur in cloud computing
        - Fixed Expenses
            - Constant costs
        - Variable Expenses
            - Costs based on usage
    - Capital expenditures and operational expenditures, or CAPEX and OPEX, are the two main categories of expenditures and costs you will hear about when discussing cloud computing
        - CAPEX
            - Expenses for long-term assets
        - OPEX
            - Ongoing costs

- **Licensing Models**
    - There are five main licensing models used in cloud computing
    - *Perpetual Licensing/Lifetime Deal (LTD)*

    - Software licensing model where the customer pays a one-time fee for a license to use the software for as long as the customer wants
    - Advantages
        - stability and certainty
        - customization and modification
    - Disadvantages
        - No support and maintenance
        - Costly upgrade
        - Costly startup
    - In the context of cloud computing, perpetual licensing can be used for cloud-based software applications
        - Customer Relationship Management systems (CRM)
        - Enterprise Resource Planning systems (ERP)
- *Subscription License*
    - Software licensing model where the customer pays a recurring fee to use the software
    - Advantages
        - Flexibility and Affordability
        - Support and Maintenance
    - Disadvantages
        - Vendor and Software lock-in

        - Costly
- *Bring Your Own License (BYOL)*
    - Licensing model in cloud computing where customers use existing software licenses in a cloud environment
    - Advantages
        - Flexibility and cost savings
        - Good option for large software license investments
    - Disadvantages
        - Compatibility
        - No support and maintenance
- *Included License*
    - Cost of the software license is included in the price of the cloud service provided by AWS
    - The included licenses model simplifies the organization's license management
- *License Manager*
    - Software tool used to manage and track software licenses
    - Advantages
        - Visibility and Control
        - Optimized software investments

        - Disadvantages
            - Complex to set up and manage
            - Require maintenance and updates

- **Instance Types and Pricing**
    - *Virtual Machine Instance*
        - Specific configuration for a virtual machine that is running within a cloud-based environment
        - Different cloud providers offer various types of instance types to accommodate different workloads
            - Small, single-core instances
                - Development and testing
            - Large, multi-core instances
                - Resource-intensive applications
        - These virtual machine instances can be monitored and reported on to understand usage patterns, resource utilization, and performance
        - Any instance will use the same hardware and receive the same performance, but the pricing will be different
    - There are various AWS instance pricing options
        - *On-demand Instances*
            - Allow to pay for compute capacity by the hour or the second with no long-term commitments or upfront payments

- On-demand instances are ideal for applications with short-term, sporadic workloads that cannot tolerate interruptions

- *Reserved Instances*
    - Offers discount of up to 75% and provides a capacity reservation when used in a specific Availability Zone
    - AWS offers three Reserved Instance types
        - *Standard Reserved Instances*
            - Offer the most significant discount and are ideal for applications with steady-state usage
        - *Convertible Reserved Instances*
            - Offer a lower discount but provide the flexibility to change the attributes of the Reserved Instance during its term
        - *Scheduled Reserved Instances*
            - Available to launch within a reserved time window
    - AWS' goal is to create predictability in demand to optimize the underlying hardware and costs
- *Spot Instances*
    - Allow to bid on spare Amazon EC2 computing capacity at a 90% discount off the on-demand price, the highest discount
    - Spot instances are excellent for large-scale, parallel batch jobs and supplements on-demand instances with lower-cost excess capacity

        - Savings Plans
            - Offer up to 72% savings on AWS compute usage by selecting a term commitment for the instances
            - The two types of Savings Plans
                - Compute Savings Plans
                    - Offer the most flexibility and can be used to reduce costs by up to 66%
                - EC2 Instance Savings Plans
                    - Offer the deepest discounts and automatically apply to specific instance families within a region
        - *Dedicated Instances*
            - Run in a Virtual Private Cloud on AWS hardware that is dedicated to a single customer or organization
            - Dedicated instances are ideal for workloads that require EC2 instances to be separated from other instances owned by other AWS accounts
        - *Dedicated Host*
            - Refers to a specific EC2 server set aside only for an organization
        - *Capacity Reservations*
            - On-demand capacity reservations reserve capacity for Amazon EC2 instances in a specific Availability Zone for any duration of time

- **Data Transfer Charges**

- *Incoming Data Transfer*
  - The transfer of data into an AWS service from a source that could be located on-premise or in another AWS service
  - All data transferred into AWS is free of charge
  - All traffic originating from the Internet that is destined to the AWS Cloud is allowed to enter for free
- *Outgoing Data Transfer*
  - The transfer of data from an AWS service to a destination that could be located on-premise or in another AWS service
  - AWS charges for outgoing data transfer on a tiered basis which means that the cost actually goes down as your usage increases
  - Conducting an outbound data transfer between two AWS services in the same region and the same availability zone is considered non-chargeable
  - These cross-availability zone data transfers are charged at the lowest level of any outbound data transfer type
- *Inter-region Data Transfer*
  - Transferring data between two different AWS regions, and this would be a chargeable event

        - It's important to consider these costs when designing an AWS architecture, especially if the application involves significant data transfer between regions
    - While all incoming data transfers are free, costs can be incurred for outgoing data transfers
    - Free outbound data transfer between two AWS services hosted within the same region and the same availability zone

- **Storage Options and Tiers**
    - Some of the different storage options available within Amazon Web Services
    - Object Storage
        - Scalable data storage solution that treats data as distinct units, or "objects", each with a unique identifier, rather than organizing data in file hierarchies or blocks
        - Amazon S3 is a highly scalable object storage service that offers a variety of storage classes to meet different needs
        - There are multiple S3 service categories available
            - S3 Standard
                - Designed for frequently accessed data, offering low latency and high throughput, but this is also one of the most expensive storage options
            - S3 Intelligent-Tiering

        - Adapts to changing access patterns, moving data between frequent and infrequent access tiers
    - S3 Standard-IA
        - Used for data that is less frequently accessed but still needs to be available quickly
    - S3 One Zone-IA
        - Best used for backup data stored in a single location, suitable for data that can be recreated if lost
    - S3 Glacier Instant Retrieval
        - Long-term storage with data access possible within minutes, and this is more expensive than the other Glacier options
    - S3 Glacier Flexible Retrieval
        - Offers more cost-effective storage but may require a few hours for data access
    - S3 Glacier Deep Archive
        - The most affordable long-term option, but accessing stored data can take up to 12 hours
    - S3 Outposts
        - Allows for the storage of data in on-premise environments within an organization's own data center
- Block Storage
    - Divides data into fixed-sized blocks and stores them with unique identifiers, commonly used for database storage and virtual machine file systems

    - Amazon Elastic Block Store
        - Block storage service that provides high performance and durability for applications that require consistent, low-latency access to data
    - Block storage is usually priced based on whether the underlying storage is being provided by a solid-state device or a hard disk drive
        - The block storage is also priced based on whether the block storage is a general purpose or provisioned "IOPS" device
    - Provisioned IOPS Device
        - Provides increased I/O speeds over a traditional general-purpose block storage device
- File Storage
    - Organizes and accesses data as a hierarchy of files and folders, much like the file system on a standard computer
    - Amazon Elastic File System (EFS)
        - File storage service that provides a simple, scalable, and secure way to share files between Amazon EC2 instances
        - EFS usage is billed based on the amount of storage used
    - The one-zone storage classes are charged at a rate of about half of the standard class rate
    - Amazon FSx service

- Amazon FSx comes in two varieties
  - Amazon FSx for Windows File Server
    - Fully managed Windows file server that provides a familiar and consistent experience for Windows applications
  - Amazon FSx for Lustre
    - Fully managed, high-performance file system that is optimized for parallel processing workloads
- The Amazon FSx service charges its customers for the amount of storage, the throughput capacity, and any backups the organization requires

- The right storage option and tier for a particular workload will depend on factors such as
  - Data access patterns
  - Data retention requirements
  - Budget
  - Type of data for storage

- **Rightsizing**
  - *Rightsizing*

        - Adjusting the infrastructure's size and capacity to match the current needs to reduce unnecessary costs and improve the performance
            - Analyzing the current AWS resource utilization
            - Identifying underutilized resources
            - Resizing the infrastructure
    - AWS Cost Explorer
        - Using machine learning algorithms can analyze the usage patterns and provide recommendations on how it can reduce costs by modifying the AWS resource configurations
    - The goal of rightsizing is to align the performance and capacity of the resources with the business needs and budget
        - Rightsizing can also improve the performance of the applications by ensuring that the resources are appropriately sized
    - To perform rightsizing
        - Collection and analysis of data regarding the AWS resource utilization
        - Analyze the data to identify any underutilized or overutilized resources
        - Implement those changes
    - Ultimately, we use four main strategies when rightsizing our AWS environments
        - *Scaling Down*
            - Reducing the size of a resource or the number of instances being used if those resources are consistently being underutilized
        - *Scaling Up*

            - Increasing the size or number of resources used in order to improve the performance
        - *Consolidation*
            - Merging several smaller resources into a single resource to save costs
    - Rightsizing Storage
        - Rightsize the storage resources to better match the actual utilization and decrease the costs over time
    - Rightsizing in the AWS environment entails optimizing the resource allocation based on the specific needs
        - Automation is key to ensuring that the rightsizing process isn’t just a one-time thing
    - *AWS CloudFormation*
        - Allows users to define and provision AWS infrastructure using templates, which can be designed to ensure that resources are provisioned with optimal configurations right from the start
        - Integrate the CloudFormation tool with automation tools
    - Rightsizing is a powerful strategy for optimizing the AWS environment

- **Managed Services**
    - *Managed Services*

        - AWS's offering where the cloud provider handles the operational tasks of maintaining, automating, and scaling cloud resources
    - AWS offers a variety of managed services designed to reduce the administrative burden, complexity, and costs associated with managing your own infrastructure
        - *Amazon Relational Database Service*
            - Managed relational database service that provides cost-efficient and resizable capacity while automating time-consuming administration tasks
            - This service supports several database engines
                - MySQL
                - PostgreSQL
                - MariaDB
                - Oracle
                - SQL Server
                - Amazon Aurora
            - RDS offers multiple instance types and pricing options
                - Also, it provides automated backups, Multi Availability Zone deployments, and Read Replicas
        - *Amazon DynamoDB*
            - Fully managed NoSQL database service that provides fast and predictable performance with seamless scalability
            - The DynamoDB service can significantly reduce costs related to database administration
            - DynamoDB's on-demand capacity mode offers flexible pricing
        - *Amazon Elastic Container Service*

      - Highly scalable, high-performance container orchestration service that supports Docker containers and allows to easily run and scale containerized applications on AWS
      - Amazon Elastic Container Service can help reduce costs by efficiently scheduling containers, allowing to maximize resource utilization
      - Amazon Elastic Container Service reduces administrative overhead for the organization
      - Primarily suited for users who want a more integrated and simple experience directly with AWS services
    - *Amazon Elastic Kubernetes Service*
      - Fully managed service that makes it easy to deploy, manage, and scale containerized applications using Kubernetes
      - Amazon Elastic Kubernetes Service can help reduce costs by managing the complexity of the Kubernetes infrastructure
      - Through the Elastic Kubernetes Service, one pays solely for the AWS resources utilized
      - Excels in scenarios where there's a need for more extensive orchestration capabilities and portability across multiple cloud providers

- **AWS Organizations**
  - *AWS Organizations*
    - Powerful tool that allows users to manage and govern multiple AWS accounts centrally

        - AWS Organizations tool can create an organizational structure that mirrors the company's departmental structure
            - *Organizational Unit (OU)*
                - A group of AWS accounts within an organization that it can be organized in a way that aligns with the business requirements
        - *Service Control Policies*
            - Permission boundaries to govern how various AWS services are used across the organization and all of its departments
    - *Consolidated Billing*
        - Receive a single bill for all the accounts in the organization which makes it easier to track the associated cloud computing costs
    - Amazon web services also provides cost allocation tags that can be used to categorize and track your AWS costs
        - When you activate these tags, AWS includes them in the AWS Cost and Usage Report, AWS Budgets, and AWS Cost Explorer
        - AWS Organizations can also help to manage the compliance across multiple accounts
    - AWS Organizations tool is a robust tool for managing multiple AWS accounts across different departments

- **Cost Allocation Tags**
    - *Cost Allocation Tags*

    - Provide a way to categorize AWS resources to understand which department, project, or application is spending valuable resources within a complex cloud-based environment
    - Each tag is a simple label consisting of a user-defined key and an optional value
    - AWS uses these tags to organize the resource costs on the cost allocation report
- *Chargebacks*
    - Method of allocating cloud computing costs to specific departments or projects within an organization
    - Chargebacks are usually implemented by setting up cost centers or departments within the cloud provider's billing system
    - In AWS, chargebacks are implemented using the AWS Cost and Usage Report
- Cost allocation tags can help identify high-cost resources and provide with insights that can drive cost optimization efforts
    - In an AWS Organization, a tagging approach that encompasses labels for every department or project becomes a viable option
- AWS Cost Allocation Tags are a powerful tool for managing and optimizing the AWS costs

- **Billing, Budgeting, and Cost Management Tools**
  - Amazon Web Services provides a suite of tools that are designed to help businesses manage their cloud costs effectively
  - To help with our billing, budgeting, and cost management, AWS provides the following tools
    - *AWS Budgets*
      - This tool is particularly valuable for keeping track of costs and preventing unexpected charges
      - AWS Budgets also provides a feature called "Budget Actions" that allows to define actions that AWS will automatically take if the user-defined cost thresholds are breached
      - AWS Budgets also provides budget performance history that can enable to track the usage or costs against the budgeted amount
    - *AWS Cost Explorer*
      - Graphical tool that provides insights into the user cost and usage pattern
      - The Cost Explorer's features extend beyond just providing data visualizations, though
        - This tool offers Reserved Instance Utilization and Coverage Reports that can provide insights into your reserved instance usage and can enable the optimization of your reserved instance purchases to help save you money
    - *AWS Billing Conductor*
      - Enables to customize the categorization and allocation of the AWS costs based on the unique business needs
    - *AWS Pricing Calculator*

      - Enables to estimate the cost of AWS services based on the projected usage
      - The AWS Pricing Calculator supports a wide range of AWS services, including EC2, RDS, S3, Lambda, VPC, and many others
      - One of the key benefits of the AWS Pricing Calculator is that it allows to model different scenarios and compare their costs
    - *AWS Cost and Usage Report*
      - It provides detailed information about the AWS costs and contains additional details about the usage
      - Cost and Usage Report is delivered to an Amazon S3 bucket that the client specify, and from there, it can analyze it using various AWS services
      - One of the key benefits is that it allows to perform more granular cost analysis
  - Amazon Web Services provides a suite of tools designed to help businesses manage their cloud costs effectively

- **Using Cost Allocation in AWS**
  - *Cost Allocation*
    - The process of tracking and assigning costs to different departments, business units, or projects
  - AWS provides a number of tools and features to help you with cost allocation
    - Cost Allocation Tags
      - Cost allocation tags are the foundation of AWS cost allocation

        - The tag key is a unique identifier for the tag, and the tag value is the data associated with the tag
      - Cost Categories
      - AWS Cost Explorer
      - AWS Cost and Usage Reports
    - For best results with tagging and reporting
      - Start tagging the resources as soon as possible
      - Always use consistent tag keys and values across the organization
      - Review the costs regularly
      - Always use the AWS Cost Explorer and AWS Cost and Usage Reports

Connecting to AWS

Objective 3.1 - Define methods of deploying and operating in the AWS Cloud.

- **AWS Management Console**
  - *AWS Management Console*
    - Web-based interface designed to facilitate the management and monitoring of Amazon Web Services (AWS) resources
    - The AWS Management Console is essentially the user's gateway to the vast array of services offered by AWS
    - The console provides access to nearly all AWS services
      - Amazon EC2
      - Amazon S3
      - Amazon RDS
      - Amazon DynamoDB
      - AWS Lambda
  - The AWS Management Console includes features like Resource Groups and Tag Editors that allow to organize and manage AWS resources effectively
  - The console also provides a Service Health Dashboard that offers real-time updates on the status of AWS services
  - Integrates with the *AWS Marketplace*
    - Digital catalog with thousands of software listings from independent software vendors

- This management console also has increased security features to protect your account
    - Multi-Factor Authentication (MFA)
    - Identity and Access Management (IAM)
- The console provides tools for tracking AWS costs and usage

- **Programmatic Access**
    - *Programmatic Access*
        - The ability to interact with AWS services using software application
        - The term programmatic access in AWS refers to the ability to interact with AWS services using some kind of software application
    - AWS services interact with software through
        - Application Programming Interfaces (APIs)
            - Set of rules that allow programs to talk to each other
            - In AWS, each service has its own API that defines how it interacts with other software
            - AWS APIs are RESTful and use standard HTTP methods like GET, POST, PUT, DELETE, and other commands
        - Software Development Kits (SDKs)
            - Collection of software tools and libraries that developers use to create applications for specific software packages, software frameworks, hardware platforms, and operating systems

    - SDKs uses APIs by providing pre-written code libraries in various programming languages
    - AWS has provided SDKs for a wide range of programming languages including Java, .NET, Node.js, Python, PHP, Ruby, Go, and C++
    - AWS SDKs provide a convenient way to interact with AWS services by providing libraries in some programming languages
    - Streamlines process of working with AWS services and allows developers to write fewer lines of code, which makes their code easier to read and maintain
- AWS Command Line Interface (AWS CLI)
    - Text-based interface used to interact with software and operating systems
        - Provides consistent interface for interacting with AWS services
        - Supports multiple platforms (Windows, Linux, and macOS)
        - Provides commands for a broad set of AWS services
        - Controls multiple AWS services and implements sequence of operations
- AWS Tools for Powershell
    - Set of modules that are implemented as cmdlets that provides developers and administrators with a consistent way to manage AWS services from PowerShell scripting environment

- The AWS Tools for PowerShell is a set of PowerShell cmdlets that are built with functionality exposed by the AWS SDK for. NET
- Enables to script operations on AWS resources from the PowerShell command line for easier automation of tasks and workflows

- AWS CloudFormation
  - Helps developers and administrators model and set up Amazon Web Services resources in less time by managing those resources
  - AWS CloudFormation service allows to use programming languages or a simple text file to model and provision, in an automated and secure manner
  - CloudFormation is used for deploying complex applications and automating operational tasks
  - CloudFormation is used to create and manage a collection of resources as a single unit (stack)

- **Infrastructure as Code (IaC)**
  - *Infrastructure as Code (IaC)*
    - Key aspect of provisioning that provides the capability for automating deployments
    - IaC approach places configurations in machine-readable files
  - Advantages of IaC over manual server configurations
    - Fast and consistent deployments

    - Reduces risk of errors
    - Improves collaboration
    - Versioned and auditable configurations
- IaC templates can be used to simplify deployments
    - Templates are a way of codifying infrastructure configurations and simplifying the deployment process
    - Templates allow for
        - Settings customization
        - Standardize configuration process
        - Versioned, auditable, rolled back (if necessary)
- In AWS, IaC is often implemented using services like
    - AWS Cloud Formation
    - AWS OpsWorks
    - Third-party tools like Terraform
- Infrastructure as Code has many uses
    - IaC enables to quickly and consistently create and update infrastructure across different stages of the application lifecycle
    - IaC ensures that testing, staging, and production environments are identical

        - IaC helps to eliminate inconsistencies and operational overhead that can arise from using manual process or approach
        - IaC files can be version-controlled in the same way as application code
        - IaC can also be used to automate the setup of complex infrastructures
        - Another great use of IaC is in the world of business continuity and disaster recovery
    - Infrastructure as Code plays a crucial role in disaster recovery and a key enabler of Continuous Integration and Continuous Deployment (CI/CD)

- **Connectivity Options for AWS**
    - AWS offers multiple options for establishing network connections to your AWS resources
    - *Virtual Private Network (VPN)*
        - Connection technology that establishes a secure and encrypted connection over a less secure network (Internet)
        - In AWS, there are two types of VPN connections
            - AWS Site-to-Site VPN
                - Enables to establish a secure and private tunnel from a network or on-premise data center to Amazon Virtual Private Cloud (VPC)
                - The Site-to-Site VPN connection consists of two main components

        - *Customer Gateway* is the client's side of the VPN connection and it includes on-premise routers and internal network
        - *Virtual Private Gateway* is the VPN concentrator on the Amazon side of the Site-to-Site VPN connection
      - Site-to-site VPN creates a single VPN tunnel to connect the entire network or building back into the AWS cloud over a single connection
    - *AWS Client VPN*
      - Managed client-based VPN service that enables users to securely access AWS resources from their on-premise network
      - Split-tunnel configuration allows users to select which traffic is going to be sent over the VPN connection
      - AWS Client VPN supports IPv4 and IPv6 network traffic and is integrated with AWS Identity and Access Management (IAM) service to grant access to AWS cloud environment
- *AWS Direct Connect*
  - Network service that provides an alternative to using the Internet by utilizing AWS services to establish a dedicated network connection between the network and one of the AWS DirectConnect locations
    - Bypass public internet connection
    - Useful for high-volume data transfer tasks or sensitive workloads
    - Support higher data speeds (802.1q VLANs)

- *Secure Shell (SSH)*
  - Cryptographic network protocol for operating network services securely over an unsecured network
  - SSH ensures that all traffic between the local machine and the EC2 instance is encrypted
- Public Internet
  - Used to provide end user access to the cloud-based applications and services

Identity and Access Management

Objective 2.3 - Identify AWS access management capabilities.

- **The Root User**
  - *Root User*
    - Central figure in the management and administration of the organization's AWS account
    - The root user has unrestricted access to all AWS services and resources
  - Given the root user's extensive access, it is critical that you protect this account in order to maintain the integrity and security of your AWS resources
    - To protect your root user account
      - Use a robust password
      - Enable Multi-Factor-Authentication (MFA)
      - Secure the email account associated with the root user’s account
  - Detection and monitoring are also important to configure in order to detect any attempts to break into the root user's account
    - Use the AWS CloudTrail tool to monitor and log account activity
    - Use the principles of least privilege when creating the cloud-based architectures
      - *Least Privilege*

        - Computer security concept where a user is provided the minimum levels of access necessary to perform their job functions
    - Powers of the root user
        - The root user is the only user who has the ability to terminate and close the AWS account on behalf of the organization
        - The root user is also the only user who can create, view, or delete CloudFront key pairs
        - The root user is the only one who can delete a service-linked role if the linked service allows it
    - Due to the extensive access provided to the root user account, it is imperative to safeguard the root user

- **Users and Groups**
    - User and group management is primarily handled through the AWS Identity and Access Management service
    - *Identity and Access Management (IAM) Service*
        - Allows to create and manage AWS users and groups and use permissions to allow and deny their access to AWS resources and to enforce the principle of least privilege
    - *IAM User*
        - A user in AWS consists of a name, a password and up to two access keys

        - When you create an IAM user, the user is not granted permission to access AWS resources by default
    - *IAM Group*
        - An IAM group is simply a collection of IAM users
    - When a user is added to a group, the user automatically is assigned the permissions that are attached to that group
        - Always apply the principle of least privilege when working with permissions and access rights in IAM
    - Best practices to consider when working with users and groups
        - Always create individual IAM users for each person who needs to use the AWS services and applications
        - Assign permissions to groups whenever possible
        - Always use the principle of least privilege
        - Use conditions for fine-grained control of the user's permissions
        - Regularly review and rotate credentials
    - IAM users can be assigned to multiple groups, and each group can contain multiple users

- **Policies**
    - *IAM Policies*

        - Documents that act as a formal set of permissions to define who has what kind of access to each specific resource in AWS
    - IAM policy is written in JSON format
        - JSON (Javascript Object Notation)
            - Specially formatted text file that contains key-value pairs that specify what permissions are being provided
    - Policies are attached to IAM identities, which could be a user, group, or role
    - IAM policies play an important role in implementing the principle of least privilege
    - *Managed Policies*
        - Standalone policies that can be attach to multiple users, groups, and roles in the AWS account
        - *AWS-managed Policies*
            - Created and managed by AWS and cover common use cases
        - *Customer-managed Policies*
            - Managed policies that the user create and manage in the AWS account
    - *Inline Policies*
        - Policies that the user create and manage, and that are embedded directly into a single user, group, or role
    - When creating a policy, always start with what is explicitly allowed, and then what is explicitly denied

- IAM policies are a powerful tool for managing access to the organization's AWS resources

- **Authentication Methods in AWS**
  - AWS provides a robust set of features and services for managing authentication and identity, and for enabling secure access to resources
    - These different authentication methods include
      - Traditional username and password for authentication
      - Use of access keys for programmatic access
      - Enablement of Multi-Factor Authentication (MFA)
      - Usage of Single Sign-on (SSO)
      - Use of cross-account IAM roles
  - *Identity Management*
    - The process of creating, managing, and securing user identities
  - *Authentication*
    - The process of verifying the identity of a user or application
  - Identity and Access Management (IAM)
    - Allows creation and management of AWS users and groups, as well as the ability to use permissions to allow and deny their access to AWS resources
  - Various methods of authentication used in a cloud-based environment
    - Username and Password

    - Involves verifying a user's identity by confirming they know specific credentials
- *Access Keys*
    - AWS relies on access keys that are comprised of two parts: an access key ID and a secret access key
    - Ensure that secure access is taking place when using API calls, AWS Command Line Interface and AWS Software Development Kits
- *Multi-Factor Authentication (MFA)*
    - Security system feature that mandates that users must provide two or more distinct forms of identification
    - Two-Factor Authentication (2FA)
        - Works just like Multi-Factor Authentication, but it only supports 2 factors
- *Single Sign-on (SSO)*
    - Permits a user to enter one set of credentials and build access, multiple applications, services, or websites using that same set of credentials
    - IAM supports identity federation
        - *Identity Federation*
            - Allows granting external identities, such as users from corporate directory, different permissions to access AWS resources without creating unique IAM user accounts
- Cross-account IAM Roles

            - Provide a robust feature that is used to enhance collaboration across multiple accounts within the AWS cloud
    - AWS also offers tools to help monitor and analyze the permissions set in place
        - IAM Access Analyzer
            - Helps to ensure that the AWS resources are only accessible to intended individuals or services
    - As organizations grow and evolve, the complexity of the system's permissions can inadvertently expose sensitive resources
    - Authentication and identity management are critical components in AWS

- **AWS Identity and Access Management (IAM)**
    - *Identity and Access Management (IAM)*
        - Web service that helps to securely control access to the organization's AWS resources
    - Identity and Access Management provides two functions in terms of cloud security
        - Authentication: Verifies if someone or something is who they claim to be
        - Authorization: IAM determines what actions or resources the authenticated entity is allowed to access and interact with
    - To configure identity and access management in the cloud, use the AWS Identity and Access Management service
    - In addition to setting up users for each individual service, you can also perform federation for your users and their permissions

        - Enable identity federation to allow existing identities in the enterprise to access the AWS Management Console, call AWS APIs, and access resources
        - You can also create a larger federation in AWS by linking to a non-AWS directory, like your company's own Active Directory on its Windows servers, to your own AWS services in the cloud
    - IAM Dashboard is the centralized location to manage users, groups, roles, and policies
        - Has two main sections
            - Access Management
                - used to manage users, groups, roles, and policies
            - Access Report
                - provides insights into who has access to your system
    - AWS Identity and Access Management (IAM) is the backbone of security and access control within the AWS environment
    - Through IAM, the user can ensure that the right individuals have the right access and that the cloud resources remain protected

- **AWS IAM Identity Center**
    - *AWS IAM Identity Center*
        - Offers a robust mechanism for securely establishing or linking the workforce identities while centrally orchestrating their access across AWS accounts and associated applications

        - Choices for managing
            - Construct and supervise user identities within the AWS framework
            - Mesh with the pre-existing identity source
    - The IAM Identity Center provides compatibility with a rich selection of identity sources, including but not limited to
        - Microsoft Active Directory
        - Okta
        - Ping Identity
        - JumpCloud
        - Google Workspace
        - Azure Active Directory
    - Uses
        - The IAM Identity Center is also used to grant multi-account access across the AWS domains
        - The IAM Identity Center paves the way for uninterrupted single sign-on inbound connections to a wide variety of AWS applications
        - The IAM Identity Center also grants single sign-on access to Amazon EC2 Windows instances

        - Centralized control over access to the EC2 Windows instances is now feasible, scalable, and manageable over multiple AWS accounts
        - The IAM Identity Center speeds up the setup and configuration of single sign-on access to SAML 2.0-compliant cloud applications
    - The AWS IAM Identity Center is truly a holistic solution that provides a fine-tuned identity and access governance system

- **Credential Storage**
    - AWS offers services like the AWS Secrets Manager to ensure secure storage, access, and management of these sensitive pieces of information
    - *Credentials*
        - Set of information that proves a user's or system's identity
    - *AWS Secrets Manager*
        - Aids in the management and protection of access keys, passwords, and other secrets
        - When a secret is retrieved, the Secrets Manager decrypts it and transmits it securely over an encrypted transport layer security tunnel to your local environment
        - The AWS Secrets Manager allows for the rotation of secrets on a schedule or on demand

        - The AWS Secrets Manager tool can also be used to automatically replicate your secrets to multiple AWS Regions
    - *AWS Systems Manager*
        - Offers a unified interface that enables the monitoring and management of AWS resources and on-premise infrastructure
        - AWS Systems Manager Parameter Store
            - Offers hierarchical storage for configuration data, secrets, and database strings, which can be used in scripts, code, and AWS CLI
        - The AWS Systems Manager has several other key features
            - Explorer
                - Provides key insights and analysis into the operational health and performance of AWS environment
            - OpsCenter
                - Provides a central location where operations engineers and IT professionals can view, investigate, and resolve operational issues
            - Incident Manager
                - Enables faster resolution of critical application availability and performance issues
            - Application Manager
                - Helps to investigate and remediate issues with resources in the context of the applications
            - AppConfig

    - Helps to deploy application configuration in a managed and monitored way
- Change Manager
    - Simplifies the way you request, approve, implement, and report on operational changes to the application configurations and infrastructure
- Automation
    - Allows to safely automate common and repetitive IT operations and management tasks
- Maintenance Window
    - Allows to schedule windows of time to run administrative and maintenance tasks across all instances
- Fleet Manager
    - Streamlines the remote management process for servers and edge devices
- Compliance
    - Used to automatically aggregate and display the operational data for each resource group through a common dashboard
- Run Command
    - Provides a safe and secure method of remote management for instances at scale without the need for logging into the servers
- State Manager

-
    -
        - Provides configuration management, which helps to maintain consistent configuration of the Amazon EC2 or on-premise instances
    - Patch Manager
        - Helps to select and deploy operating system and software patches automatically across large groups of instances
    - Distributor
        - Helps to securely distribute and install software packages, such as software agents
- Both the AWS Secrets Manager and AWS Systems Manager Parameter Store are essential tools for safeguarding the credentials in the cloud

Compute Services

Objectives:

- 3.3 – Identify AWS compute services.
- 3.8 - Identify services from other in-scope AWS service categories.

- **Elastic Cloud Compute (EC2)**
  - *Amazon Elastic Cloud Compute (EC2)*
    - Core component of the AWS platform that provides scalable computing capacity in the cloud
    - Instances are rented to the organization based on the different instance types available
  - The AWS Elastic Compute Cloud has a variety of different instance types or categories
    - *General-purpose Instances*
      - Designed to provide a balance of compute, memory, and networking resources
      - Under general-purpose instances, you will find various types under this category
        - T2
        - T3
        - T3a
        - T4g
      - M5 Instance

        - Provides with a good balance of compute, memory, and networking (T-series)
    - *Compute-optimized Instances*
        - Designed for compute-bound applications that require high-performance processors
    - *Memory-optimized Instances*
        - Ideal for memory-intensive applications that require a high memory-to-CPU ratio
    - *Storage-optimized Instances*
        - Designed for workloads that require high, sequential read and write access to large datasets on local storage
    - *High-performance Computing Instances*
        - Designed to provide the best price-performance for running High-Performance Computing workloads at scale on AWS
        - High-performance computing instances are engineered to handle large, complex simulations, deep learning workloads, and other computational-intensive tasks
    - *Accelerated Computing Instances*
        - Uses hardware accelerators, or co-processors, to perform some functions efficiently in software running on general-purpose CPUs
        - Accelerated Computing Instances are designed to reduce the time to process data-intensive tasks that can benefit from hardware acceleration
        - Accelerated Computing Instances can offload compute-intensive portions of the application onto the GPU or co-processor

- **Auto-scaling with EC2**
    - *Auto-scaling*
        - Dynamic adjustment of computational resources based on the actual demand or load at any given time
        - Auto-scaling allows systems to automatically increase or decrease the number of resources
        - Utilize the auto-scaling service within EC2 to maintain the desired number of instances and adjust the number of instances, upwards or downwards, in real time
        - Auto Scaling is effective in applications that experience varying levels of demand at different times
    - To set up auto-scaling
        - Launch Configuration
            - Set up and use a launch configuration template to launch EC2 instances using auto-scaling capability
        - Auto Scaling Groups
            - Uses the specified launch configuration to launch instances when certain conditions are met
            - The Autoscaling group is used to specify the details
                - Group Name
                - Initial Instance Count
                - Network

        - Creating collections of EC2 instances (Auto Scaling Groups) can specify the minimum number of instances in each Auto Scaling Group
        - Elasticity is the ability to scale computing resources up or down in response to changing demand
    - Scaling Policies
        - Used to determine when Auto Scaling should launch or terminate instances
        - Auto Scaling system can automatically launch more instances to accommodate increased demand using 'scaling out'
        - Auto Scaling system can automatically stop or terminate instances that are no longer needed to reduce cloud computing costs in a process known as 'scaling in'
    - Health Checks
        - Used to determine the status of the instances in the Auto Scaling group
        - Use EC2 status checks or Elastic Load Balancer health checks to configure health checks
        - Use multiple Availability Zones to configure Auto Scaling
    - Configure Optional Notifications
        - Can be set up to send notifications to launch or terminate an instance automatically
- *Amazon Simple Notification Service (SNS)*

        - Communication channel to which messages can be sent by publishers and seen by subscribers

- **Load Balancing with EC2**
    - *Load Balancing*
        - Practice of distributing incoming network traffic across multiple servers to ensure no single server is overwhelmed with too much demand
        - Load balancing is used to help maximize the throughput, reduce the latency, and ensure fault tolerance within our applications
        - Load balancing works in the AWS cloud and helps to achieve a high-performance experience for users
    - Elastic Load Balancing (ELB) Service
        - Used to automatically disseminate incoming application traffic across several EC2 instances
            - Provides higher levels of fault tolerance
            - Identifies unhealthy instance
            - Enables elastic load balancing service
            - Used in Amazon VPC to distribute traffic
    - There are three different types of load balancers in AWS
        - *Application Load Balancer*

            - Suited for load balancing of HTTP and HTTPS traffic that provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers
                - Routes traffic to Amazon VPC
                - Operates at Layer 7 (OSI model)
        - *Network Load Balancer*
            - Used for load balancing of network traffic using TCP, UDP, and TLS
                - Suited for high-performance traffic
                - Routes traffic to Amazon VPC
                - Operates at Layer 4 (OSI model)
        - *Gateway Load Balancer*
            - Used to deploy, scale, and run third-party virtual networking appliances
                - Customers deploy appliances from the preferred vendor
                - Operates at Layers 3 & 4
    - Elastic Load Balancing (ELB)
        - Managed service with the AWS Cloud that allows to focus on delivering applications and not installing fleets of load balancers across cloud-based networks
        - When using the Amazon VPC, the elastic load balancing service can provide robust security features
        - The Elastic Load Balancing service is used with other AWS services such as EC2, ECS, EKS, and tools like Amazon CloudFormation and the AWS Billing

        - The Elastic Load Balancing can support load-balancing capabilities
        - Elastic Load Balancing in Amazon EC2 automatically distributes incoming application traffic across multiple EC2 instances
        - The elastic load balancing service can support a variety of load balancers, each designed for specific scenarios and traffic types
            - Application Load Balancer
                - HTTP and HTTPS
            - Network Load Balancer
                - TCP, UDP, and TLS
            - Gateway Load Balancer
                - Deploy, scale, and run third-party virtual networking appliances

- **Containerized Compute**
    - Containerized Compute
        - Way of running applications in isolated containers
    - Containerized compute provides us with four main benefits
        - *Portability*
            - Easily moved from one environment to another
        - *Scalability*
            - Easily scaled up or down in response to fluctuating demands
        - *Efficiency*
            - Shares the same OS

        - *Security*
            - Each container is isolated from each other
    - There are some drawbacks to using containerized compute instead of traditional virtual machines or EC2 instances
        - Complexity
            - Containerized applications can be more complex to manage than traditional applications, so this can lead to additional labor costs since they require a more skilled workforce to operate
        - *Vendor Lock-in*
            - Occurs when an organization becomes dependent on a particular vendor's products or services
    - 3 main types of containerized compute used in AWS
        - *Amazon Elastic Container Service (ECS)*
            - Fully managed container orchestration service that makes it easy to run, manage, and scale containerized applications
            - Elastic Container Service is scalable and reliable, and a cross-platform container service
            - Elastic Container Service is considered to be proprietary technology developed by AWS
    - To use ECS
        - You would first need to create an ECS cluster
            - *Cluster*
                - the group of EC2 instances that are used to run the containers

        - Once you have created a cluster, you can create a task definition
            - *Task Definition*
                - a blueprint for a container
        - Once you have created a task definition, you can create a service
            - *Service*
                - the group of tasks that runs in application
            - When you create a service, you specify the task definition that you want to use, the number of tasks that you want to run, and the deployment strategy
    - Amazon Elastic Kubernetes Service (EKS)
        - Managed Kubernetes service that makes it easy to run Kubernetes on

          AWS
        - *Kubernetes*
            - Open-source container orchestration system that allows to deploy, manage, and scale containerized applications
        - Advantages
            - Scalable and flexible
            - On-premise/hybrid clouds/multiple cloud environments
        - To use EKS
            - You would first need to create an EKS cluster
                - *Cluster*

      - the group of EC2 instances that are used to run the containers
    - Once you have created a cluster, you can create a Kubernetes manifest
      - *Kubernetes Manifest*
        - the blueprint for a Kubernetes cluster
    - Once you have created a Kubernetes manifest, you can deploy it to your EKS cluster
- Amazon Elastic Container Registry (ECR)
  - Fully managed Docker image registry that makes it easy to store, manage, and deploy Docker images
  - Elastic Container Registry is great for storing Docker images to be used in production, development, and testing environments
  - To use ECR
    - You would first need to create an ECR repository
      - *Repository*
        - a container for storing Docker images
    - Once you have created a repository, you can push your Docker image to the repository
      - Use the 'docker push' command to push the Docker image to the repository

- Once you have pushed your Docker image to ECR, you can use it with the Elastic Container Service or the Elastic Kubernetes Service
  - Use ECS or the EKS to specify the repository name and the image tag of the task definition or Kubernetes manifest

- Difference between the Elastic Container Service and the Elastic Kubernetes Service

Elastic Container Service (ECS)	Elastic Kubernetes Service (EKS)
Managed service	Unmanaged service
Docker container	Kubernetes container
Simple applications	Complex applications

  - Elastic Container Service (ECS)
    - Run, manage, and scale containerized applications
  - Elastic Kubernetes Service (EKS)
    - More flexibility and control
  - Elastic Container Registry (ECR)
    - Managed Docker image

- **Serverless Compute**
  - *Serverless Compute*

        - Cloud computing model in which the cloud provider does the provisioning, scaling, and management of the underlying compute resources
        - Serverless compute can provide us with some great benefits, such as cost savings, scalability, and ease of use
    - Lambda and Fargate are two of the most popular serverless compute services
    - AWS Lambda
        - Serverless compute service that runs code in response to events and automatically manages the underlying compute resources needed
        - The code executed by the AWS Lambda serverless computing service is a Lamba function
        - AWS Lambda is suitable for a wide range of applications and use cases
            - Processing HTTP requests
                - Allow to quickly and easily respond to incoming requests
            - Sending emails
            - Executing batch jobs
            - Real-Time Stream Processing
            - Extract-Transform-Load (ETL) Jobs
                - Used to move data from one data store to another
    - *AWS Fargate*
        - Designed for running containers without having to manage the underlying infrastructure

        - AWS Fargate is ideal for applications that require isolation at the task level or complex networking requirements
    - Comparison of Lambda and Fargate

AWS Lambda	AWS Fargate
Runs individual functions in response to events	Runs containers without managing clusters/services

        - AWS Fargate is ideal for applications that require isolation at the task level or complex networking requirements
        - AWS Fargate is commonly used with
            - Microservices
                - Can be packaged as a container, and each container can be scaled independently
                - Batch Processing
                    - Can be packaged as a container and run as an individual AWS Fargate task
                - Machine Learning
                    - Can be packaged as a container and run as an AWS Fargate task
        - Each microservice or component can be containerized and managed using the AWS Fargate service

- **Other Compute Services**
    - AWS offers a variety of compute services tailored to different needs and use cases
    - *Amazon LightSail*
        - Virtual private server service that offers simple and cost-effective compute instances
        - LightSail can launch a virtual machine preconfigured with SSD-based storage, DNS management, and a static IP
        - LightSail offers a variety of instance plans that bundle different levels of memory, processing power, storage capacity, and data transfer at various price points
        - Amazon LightSail is ideal for
            - Simpler workloads, quick deployments, and simple management interface
            - Hosting low-traffic websites or web applications
            - Setting up development and test environments
            - Easy way to gain experience with AWS services
    - *AWS Elastic Beanstalk*
        - Fully managed service that simplifies the deployment and scaling of web applications and services developed in Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker

- AWS Elastic Beanstalk is ideal for
  - Deploys code without the underlying infrastructure
  - Hosts web applications
  - Deploys and run applications written in a variety of languages and frameworks
  - Hosts RESTful APIs
  - Deploys individual microservices

- *AWS Batch*
  - Fully managed batch processing service that allows developers, scientists, and engineers to easily and efficiently run batch computing jobs
  - AWS Batch is ideal for workloads that are parallelizable and compute-intensive
  - AWS Batch is the training of machine learning models or the running of inference jobs

- **End-user Computing Services**
  - *End-user Computing Service*
    - Cloud-based service that provides users with access to applications, data, and devices from any location
    - Using end-user computing services, users can scale up or scale down resources on demand
  - *Amazon AppStream 2.0*

        - Fully managed application streaming service that makes it easy to deploy and manage applications for a wide range of devices
        - AppStream 2.0 can easily scale any number of users without acquiring, provisioning, or operating hardware or infrastructure
        - Using AppStream 2.0 allows to stream applications to employees' devices
    - *Amazon WorkSpaces*
        - Fully managed desktop-as-a-service (DaaS) that provides users with secure, reliable, and scalable access to virtual desktops
        - WorkSpaces also offers a variety of features that can help organizations improve security and compliance
            - Single Sign-on
            - Remote Access
            - Data Loss Prevention
        - Desktop-as-a-Service (DaaS)
            - Cloud-based service that provides users with access to virtual desktops
            - Benefits of using DaaS
                - Cost-saving
                - Increase flexibility
                - Improve security
                - Simplify IT management requirements
    - *Amazon WorkSpaces Web*

- Fully managed, secure, browser-based solution that allows to access internal websites and cloud-based applications
- WorkSpaces Web is non-persistent and does not store data after each session
- WorkSpaces Web offers a secure web browser-based productivity
- However, the prevailing methods to safeguard web browsing traffic can be either too lenient, costly, complicated, or a combination of these things

Network Services

Objective 3.5: Identify AWS network services.

- **Software Defined Networking (SDN)**
  - *Software Defined Networking (SDN)*
    - Approach to networking that uses software-based controllers or application programming interfaces (APIs) to communicate with underlying hardware infrastructure and direct traffic on a network
  - *Infrastructure as Code (IaC)*
    - Includes the provisioning of architecture in which deployment of resources is performed by scripted automation and orchestration
  - 3 portions of a typical network architecture
    - *Control Plane*
      - Responsible for carrying the traffic that provides the signals to and from a router such as those used in sharing information and building routing tables
    - *Data Plane*
      - Used to carry user traffic on the network
      - It's the data plane that's actually moving the data around the network
    - *Management Plane*
      - Used to administer the routers and switches inside of the network
  - In a traditional or conventional network, these different planes all exist in a physical piece of hardware

- With SDN, these functions are incorporated into a virtualized device or decoupled device that focuses on a single plane's functions at a time using APIs provided by the vendors to communicate between these three planes
- Using an SDN has some advantages over conventional networks, such as having the ability to mix and match products from different vendors
- SDN also provide organizations with increased choices in their network development
- The best benefit of an SDN is that it can allow for fully automated deployment of a network within the cloud
- Losing connectivity to the SDN controller could cause the entire network to go down
- The use of a centralized controller in SDNs creates a real target for attackers to focus on

- 3 Main Types of Software Defined Networks (SDN)
  - *Open SDN*
    - Open source variant of SDN that relies on open-source technology like OpenFlow, OpFlex, and OpenStack to operate
  - *Hybrid SDN*
    - Network that employs traditional SDN protocols to operate itself
  - *SDN Overlay*

            - Method of using software to create layers of network obstruction that can be used to run multiple separate and discrete virtualized network layers on top of the physical network
    - Software-Defined Network (SDN) is a way of managing and controlling networks using software instead of traditional hardware devices

- **Virtual Private Cloud (VPC)**
    - *Amazon Virtual Private Cloud (VPC)*
        - Empowers users to provision a logically isolated section of the AWS Cloud where they can launch various AWS resources in a virtual network that they have defined
        - Amazon VPC is an integral part of the larger concept known as Infrastructure as Code (also known as IAC)
        - *Infrastructure as Code (IaC)*
            - Provisioning of architecture where the deployment of resources is performed by scripted automation and orchestration
    - The Amazon Virtual Private Cloud provides us with a lot of different features and core components that you should be aware of
        - *Subnets*
            - A range within the Virtual Private Cloud that can be allocated to the instances that will be launched
        - *Route Tables*

        - Contains a set of rules, called routes, that determine where network traffic is directed by switches and routers within the virtual private cloud
    - Internet Gateways
        - A horizontally scalable, redundant, and highly available VPC component that allows communication between instances in the VPC and the Internet
    - *Network Address Translation Gateways*
        - Used to enable instances in a private subnet to connect to the Internet or other AWS services but prevent the Internet from initiating a connection with those instances
    - *Security Groups*
        - Preferred way to control inbound and outbound traffic for the instances and can be used to control traffic on a per-port and per-protocol basis
    - *Network Access Control Lists (ACLs)*
        - Can be used to supplement security groups, but they should not be used as a replacement for security groups
    - *VPC Peering*
        - A networking connection between two VPCs that enables users to route traffic between them privately
    - *VPC Endpoints*
        - Allow private connectivity to services hosted in AWS from within the VPC without using an Internet Gateway, VPN, or AWS Direct Connect

        - VPN Connections
            - VPN connections can be created between the user's VPC and remote network or between two VPCs
    - In a traditional network, these functions might occur within various physical devices
        - Routers
        - Firewalls
        - Unified Threat Management Systems
        - Switches
    - To set up an Amazon VPC, the user must define these components and the interactions between them
    - Amazon Virtual Private Cloud (VPC) is a logically isolated section of the AWS Cloud where users can launch AWS resources

- **VPC Security**
    - Network Access Control Lists (ACLs)
        - Operates at the subnet level within the Virtual Private Cloud
        - Each Network ACL is comprised of a numbered list of rules that AWS will process in ascending order
        - Network ACLs are used to create two sets of rules

        - Inbound Rules
            - Control the traffic that's allowed to enter the subnet
        - Outbound Rules
            - Govern the traffic that's allowed to leave the subnet
    - Security Groups
        - Designed to function at the instance level within the Virtual Private Cloud (VPC)
        - Each of the Security Groups is comprised of a set of defined rules that will allow inbound and outbound traffic
    - The use of both Network ACLs and Security Groups offers us a lot of different security advantages
        - Provide granular control over the network traffic
        - Tailor the security to the user's unique requirements
        - Create a robust and multi-layer defense for the VPC
    - It is critically important that the Network ACLs and Security Groups are correctly configured for security

- **DNS**
    - *Domain Name System (DNS)*
        - Used to help network clients find a website using human-readable hostnames instead of numeric IP addresses
    - *Fully Qualified Domain Name (FQDN)*

        - Complete address of an internet host or computer

    - DNS is setup as a hierarchy
        - *Root Level*
            - Highest level in the DNS hierarchy tree
            - the root name server answers requests in the root zone
        - *Top Level*
            - organizational hierarchy (.com, .net, .org, and others)
            - geographic hierarchy (.uk for the United Kingdom, .fr for France, .it for Italy, and other countries)
        - Second Level
            - These domains sit directly below the top level domain
        - Sub-Domain
            - A domain that is a part of another domain
        - Host
            - Lowest and most detailed level inside of the DNS hierarchy and refers to a specific machine
    - *Uniform Resource Locator (URL)*
        - Address of a given unique resource on the Web
            - has https:// or http:// or ftp://
            - identifies protocol
    - Different Types of DNS Records
        - Each Type holds a different type of record

        - Types
            - A
            - AAAA
            - CNAME
            - MX
            - SOA
            - PTR
            - TXT
            - SRV
            - NS
    - It is common to set up an internal DNS service that lets the cloud instances within the same network or private cloud access each other
        - To do this, internal A records are created and internal pointer (PTR) records are also created in the reverse zone
    - *External DNS*
        - Records created around the domain names that users purchase from a central authority and use on the public Internet
    - Each DNS record has a *Time-To-Live (TTL)*
        - A setting that tells the DNS resolver how long to cache a query before requesting a new one
    - DNS Resolver
        - Also known as a DNS cache

    - is located on the user's individual host
  - Lookup Types
    - *Recursive Lookup*
      - DNS server will hunt it down and report back to the user's resolver
    - *Iterative Lookup*
      - DNS resolver will continually query DNS servers until it finds the one with the IP for the domain

- **Amazon Route 53**
  - *Amazon Route 53*
    - Scalable and highly available DNS web service that can be used to connect end user requests to user's Internet applications
  - Amazon Route 53 performs three main functions
    - Domain Registration
      - When registering a domain name, the user is reserving a unique name on the Internet that can be used to identify their website or other resources
    - DNS Routing
    - Health Checking of Resources
  - Through Route 53, we can configure how different domain names and their associated DNS records are routed to our infrastructure running within the AWS Cloud

        - Amazon EC2 Instances
        - API Gateway Endpoints
        - AWS Fargate Compute Capability
        - Elastic Load Balancing Load Balancers
        - Amazon S3 Buckets
    - Amazon Route 53 is organized as a hierarchy of domains
    - *Hosted Zone*
        - Collection of DNS records for a domain
        - Container that holds information about how users want to route traffic on the Internet for a specific domain
    - One of the biggest benefits of using Amazon Route 53 is it has built-in integrations to other AWS services
    - Amazon Route 53 is built using AWS's highly available and reliable infrastructure

- **Network Edge Services**
    - *Network Edge Services*
        - Suite of AWS services that are designed to deliver content and applications closer to end-users
    - *Amazon CloudFront*

        - A content delivery network (CDN) that delivers content to users from servers that are located close to them
        - *Content Delivery Network (CDN)*
            - Globally distributed network of proxy servers and data centers that work together to provide high availability and performance by distributing services spatially relative to end-users
        - Amazon CloudFront is a fast content delivery network service offered by AWS that securely delivers data, videos, applications, and APIs
        - CloudFront is integrated with AWS both physically by its network connections and with other AWS services like
            - Amazon S3
            - Amazon EC2
            - Elastic Load Balancing
            - Amazon Route 53
    - *AWS Global Accelerator*
        - Managed networking service that improves the availability and performance of applications with local or global users by providing static IP addresses that act as a fixed entry point
        - Amazon Global Accelerator uses the AWS global network to optimize or shorten the network path

- **Amazon API Gateway**

- *Amazon API Gateway*
    - Fully managed service designed to streamline the process for developers to devise, publish, maintain, monitor, and safeguard APIs at scale
    - Amazon API Gateway facilitates the creation of RESTful APIs and WebSocket APIs
- One of the key strengths of API Gateway is its ability to handle all the tasks involved in accepting and processing a tremendous number of concurrent API calls
    - To achieve this, the API gateway must be able to conduct the following
        - Traffic Management
        - CORS Support
        - Authorization and Access Control
        - Throttling
        - Monitoring
        - API Version Management
- API Gateway features
    - has no minimum fees or startup costs
    - The API Gateway can support RESTful, REST, and WebSocket APIs
        - *RESTful APIs*
            - Designed for serverless workloads and HTTP backends that rely on using HTTP APIs
        - *REST APIs*

            - More general-purpose type of API that can be used for a variety of workloads
        - *WebSocket APIs*
            - Type of API that is designed for real-time communication
    - Advantages
        - API Gateway allows users to run multiple versions of the same API simultaneously
        - Cost savings on a scale because of the tiered pricing model it uses for its API requests
        - The API Gateway also offers easy monitoring with performance metrics and information on API calls, data latency, and error rates
    - In terms of security, the API Gateway full supports flexible security controls by authorizing access to your APIs with AWS Identity and Access Management and Amazon Cognito tools
        - API Gateway also offers embedded support for OIDC and OAuth2
        - Users can execute or code their own Lambda authorizer function from within the AWS Lambda service

- **Amazon VPN**
    - *Amazon Virtual Private Network (VPN)*
        - A managed service that enables users to securely connect their on-premise networks to AWS

    - Amazon VPN supports a variety of VPN protocols, including IPsec and OpenVPN
  - There are two types of Amazon Virtual Private Networks commonly used
    - *AWS Client VPN*
      - Managed client-based VPN service that enables users to securely access their AWS resources and networks from anywhere
      - To use AWS Client VPN
        - Setup configurations
        - Install client software
        - Establish a connection
        - Access
    - *AWS Site-to-Site VPN*
      - Connects users' on-premise networks, remote offices, or branch offices securely back to their own Amazon VPC
      - To use AWS Site-to-Site VPN
        - Set up configuration
        - Create tunneling
        - Access connection
  - Benefits
    - Fully scalable
    - Can be configured to also use a Mobile Device Management solution
      - *Mobile Device Management (MDM)*

        - Allows users to enforce their security policies by examining any device that attempts to connect to their network over the VPN
  - Users can allow Amazon VPC to host resources behind their organization's firewall

- **Amazon Direct Connect**
  - *AWS Direct Connect*
    - Cloud service that provides a more predictable and secure network experience compared to a standard Internet-based connection
    - AWS Direct Connect allows you to establish either a hosted connection facilitated by an AWS Direct Connect Delivery Partner or a dedicated connection directly from AWS
    - AWS Direct Connect service offers a wide range of speeds, starting from 50 Mbps and scaling up to 100 Gbps
  - To secure your communications, AWS Direct Connect provides multiple encryption options
    - If you are using one of the highest speed connections, such as the 10 gigabits per second and 100 gigabits per second connections
      - They utilize the IEEE 802.1AE point-to-point encryption protocol known as “MACsec” at certain AWS data centers around the world

        - If you are using a slower speed connection, or are using Direct Connect in a data center that doesn't support MACsec
            - Users can use an AWS Site-to-Site VPN to create a secure connection using “IPsec”
    - The purpose of the AWS Direct Connect service is to allow users to seamlessly connect their on-premise network to the AWS cloud

Storage Services

Objective 3.6: Identify AWS storage services.

- **Storage Features**
  - To effectively use cloud-based storage, it is important that you understand the different storage features
  - *Compression*
    - Used by cloud storage providers to reduce the amount of storage space needed to store data in a cloud-based system
    - Compression is implemented using built-in compression features in the various AWS services
      - S3 Object Compression is used for the compression of data before it is transferred into S3 to reduce transfer time and storage costs
      - Amazon Redshift uses columnar storage and automatic compression to store data more efficiently
      - AWS DataSync automatically compresses data during network transfers to optimize transfer speeds and reduce your costs
  - *Deduplication*
    - Technique used to eliminate duplicate copies of data stored in the cloud to save storage space and reduce storage costs
    - Most AWS storage services offer deduplication as a built-in feature of the service

            - Amazon S3
                - Deduplication at the object level
            - Amazon Elastic Block Store
                - Deduplication at the block level
            - Amazon Elastic File Storage & Amazon FSx service
                - Deduplication at the file level
    - *Capacity on Demand*
        - Allows cloud storage providers to offer the ability to scale storage capacity up or down as needed
        - Most of the AWS storage services provide capacity on demand as one of their key features
            - Amazon S3
                - Access a virtually unlimited amount of storage
            - Amazon Elastic Block Store
                - Modify volume size, type, and adjust the performance
            - Amazon Elastic File Storage
                - Allows to automatically scale up or down when needed

- **Storage Characteristics**
    - *Cloud Storage*
        - Type of data storage where data is stored on remote servers that can be accessed over the internet
    - 2 common cloud storage characteristics

- *Performance*
  - Characteristic that refers to the speed and reliability of data access in cloud storage
  - A cloud storage provider with high performance will be able to quickly access and transfer large amounts of data
  - High performance storage can also be more expensive
- *Hot versus Cold Storage*
  - *Hot Storage (Active Storage)*
    - Characteristic that refers to data that is frequently accessed and updated
    - Hot storage can also be more expensive than cold storage
  - *Cold Storage (Archival Storage)*
    - Characteristic that refers to data that is infrequently accessed and updated
    - Data stored in cold storage may take longer to access and retrieve

- Each cloud service provider uses different names for their different types of hot and cold storage
  - Amazon Web Services
    - Amazon Elastic Block Store (Hot Storage)
    - Amazon S3 Glacier Service (Cold Storage)
      - S3 Glacier is an excellent choice for businesses and individuals looking to store vast amounts of data for the long term
  - Amazon also uses the Amazon S3 as a form of hot storage

- Hot Storage
    - Memory or Solid State Devices
- Cold Storage
    - Hard Disk Drives or Tape Backups

- **Object Storage**
    - Object Storage/Object-based Storage
        - Type of cloud storage that stores data as individual objects
    - Amazon S3
        - Objects used to store data are called ‘buckets’
        - Designed for 99.999999999% availability and stores data for millions of companies
        - Each object can be stored as a single object up to 5TB
    - The S3 service also comes with features that enables
        - Attachment of metadata tags to objects
        - Data movement and storage across several S3 storage classes
        - Configuration of data access controls
    - Objects are stored in S3 buckets and can be organized using shared names known as 'prefixes'

        - Users can append up to 10 key-value pairs, referred to as S3 object tags, to each object
    - S3 Batch Operations are also used to simplify the management of your data at any scale and they allow you to
        - Copy objects between buckets
        - Replace object tag sets
        - Modify access controls
        - Restore archived objects
    - Amazon S3 supports data version control, accidental deletion prevention, and data replication to the same or a different AWS Region
    - Collection of different capabilities under the 'object storage' category
- Amazon S3 offers a range of storage classes designed for specific use cases and access patterns
    - S3 Standard
        - Designed for frequently accessed data, offering low latency and high throughput, and is one of the most expensive storage options since it is optimized for frequently accessed data
        - S3 Standard is ideal for various use cases
            - Cloud applications
            - Dynamic websites

        - Content distribution
        - Mobile and gaming applications
        - Big data analytics
    - Storage classes can be configured at the object level to allow a single bucket to contain objects stored across different service options
    - S3 Standard also can be used with the S3 Lifecycle management feature
        - S3 Lifecycle Management automatically transition objects between storage classes without requiring changes to the application
- S3 Intelligent-Tiering
    - Adapts to changing access patterns and moves data between frequent and infrequent access tiers
    - With S3 Intelligent Tiering, data is moved to the most cost-effective access tier without impacting performance, imposing retrieval fees, or adding operational overhead
    - S3 Intelligent-Tiering does not impose retrieval charges
    - Objects stored in the optional Deep Archive tiers need to be restored using RestoreObject before being retrieved
- S3 Standard-IA (Standard Infrequent Access)
    - Designed for less frequently accessed data that requires quick retrieval when needed

    - Combines the high durability, high throughput, and low latency of S3 Standard with a low per GB storage price and a per GB retrieval fee
    - Storage classes in S3 Standard-IA can be configured to enable a single bucket to contain objects stored across different classes
    - S3 lifecycle policies can be utilized to automatically transition objects between storage classes without modifications to existing applications
- S3 One Zone-IA (S3 One Zone-Infrequent Access)
    - Storage solution designed for less frequently accessed data that requires quick access when needed
    - This storage class is an excellent choice for customers seeking the following
        - Cost-effective solution for infrequently accessed data
        - Storing secondary backup copies
        - Cost-effective storage solution for data replication
    - For longer-term storage and for data that does not need to be accessed quickly or frequently
        - Cold Storage
            - Infrequently accessed data
            - Cold storage is less expensive than hot storage
- S3 Glacier Instant Retrieval
    - Long-term storage with data access possible within minutes

- S3 Glacier Instant Retrieval offers the quickest access to archival storage while providing the same throughput and millisecond access
- S3 Glacier Flexible Retrieval
    - Offers more cost-effective storage but may require a few hours for data access
    - It's an excellent solution for
        - Archival data
        - Retrieval options
        - Backup
        - Disaster Recovery
        - Offsite data storage needs
- S3 Glacier Deep Archive
    - Most affordable, long-term option for archival storage, and designed to support long-term retention and digital preservation of data accessed
    - This service was really designed for customers who operate in highly-regulated sectors such as
        - Financial services
        - Healthcare
        - Public sector

        - S3 Glacier Deep Archive is ideal solution for backup and disaster recovery use cases
        - S3 Glacier Deep Archive stored data can take up to 12 hours to retrieve
    - S3 Outposts
        - Allow for the storage of data in on-premise environments within an organization's own data center
        - S3 Outposts help to simplify the process of storing and retrieving data on Outpost
        - S3 Outposts offer capabilities to secure data, control access, and tag and report stored data
- AWS provides a portfolio of data transfer services to provide the right solution for any data migration project
    - Hybrid Cloud Storage
    - Online Data Transfer
    - Offline Data Transfer
- AWS Storage Gateway
    - Hybrid cloud storage service that connects and extends on-premise applications to AWS Storage
- AWS DataSync
    - Makes it easy and efficient to transfer hundreds of TBs and millions of files into Amazon S3

- AWS Data Exchange for Amazon S3
    - Used to accelerate the time to insight with direct access to a data provider's Amazon S3 data
- By default, users only have access to the S3 resources they create
    - You can grant access to other users using the following
        - AWS Identity and Access Management (IAM)
        - Access Control Lists (ACLs)
        - Bucket policies
        - S3 Access Points
        - Query String Authentication
- Amazon S3 also offers flexible security features to prevent unauthorized users from accessing your data
    - VPC Endpoints
    - Server-side Encryption
    - Client-side Encryption
    - S3 Block Public Access
- Microsoft Azure
    - Objects used to store data are called ‘blobs’
- Google Cloud Platform
    - Objects used to store the data called ‘Google Cloud Storage’

- **Block Storage**
    - *Block Storage*
        - Foundational storage architecture that divides and manages data in fixed-sized chunks, each labeled with a unique identifier
        - Block storage is great for use-cases that require rapid, consistent Input and Output operations with low latency
    - *Amazon Elastic Block Store (EBS)*
        - High-performance block storage service that was designed for use with Amazon EC2 for both throughput and transaction-intensive workloads at any scale
        - Amazon EBS volumes are placed in a specific Availability Zone where they are automatically replicated to protect users from the failure of a single component
        - EBS provides SSD-backed storage for transactional workloads
        - EBS also provides options for hard disk drive based storage which can provide the throughput needed intensive workloads
            - *Elastic Volumes*
                - Feature of Amazon EBS that allows to dynamically increase capacity, tune performance, and change the type of live volumes with no downtime or performance impact

    - Amazon Elastic Block Store also provides the ability to save point-in-time snapshots of the user volumes to the Amazon S3 buckets
    - When configuring EC2 instances, it will allow the option of using a general purpose instance or an EBS-optimized instance that provides IOPS provisioned volumes in EBS
    - Amazon EBS also offers seamless encryption of EBS data volumes, boot volumes, and snapshots
    - For higher levels of performance in the EC2 instances, enable the Multi-Attach option on an EBS Provisioned IOPS io2 or io1 volume
  - *Torn Write Prevention*
    - Feature of Amazon EBS that ensures that full 16KiB write operations are used when writing to the block storage device's volumes
  - *Instance Store*
    - Directly attached to an EC2 instance, and it offers temporary block-level storage for that EC2 instance
    - Instance stores are perfect for transient data, like buffers, caches, and other temporary content
  - Block storage is used to manage data in distinct, fixed-sized chunks called blocks
  - Amazon Elastic Block Store service is used for the creation of storage volumes which can be attached to Amazon EC2 instances

- **File Storage**
  - *File Storage*
    - Type of cloud storage that is optimized for storing and sharing files
    - File storage is often used for file sharing and collaboration
    - it can be accessed via protocols like NFS and SMB
  - AWS offers a variety of file system services optimized for different applications and use cases
  - *Amazon Elastic File System (EFS)*
    - Serverless, fully elastic file storage system designed to share file data without the need for provisioning or managing storage capacity and performance
    - Amazon EFS supports a wide range of use cases from hosting user's home directories to hosting business-critical application files, including storage for
      - Containerized and Serverless Apps
      - Big Data Analytics
      - Web Serving and Content Management
      - App Development and Testing
      - Media and Entertainment Workflows
      - Database Backups
    - Elastic File System provides an NFS-shared file system storage for Linux workloads to quickly create and configure file systems

- A fully managed file system can be created in seconds using the AWS Management Console, the AWS Command Line Interface, the AWS API, or the AWS SDK
- Amazon EFS is designed for high availability and durability, boasting a 99.999 999 999 percent durability rating, which is eleven 9's, and up to 99.99 percent availability rating
- Amazon EFS's storage capacity grows and shrinks automatically as the user add and remove files
- Amazon EFS also includes data protection features
  - EFS Replication
    - Used to replicate the file system data to another AWS Region or within the same Region in just a few steps
  - AWS Backup
    - Fully managed backup service that centrally manage and automate the backup of the Amazon EFS file systems

- *Amazon FSx*
  - Cloud-based service designed to simplify and cost-effectively launch, run, and scale feature-rich, high-performance file systems
  - Amazon FSx is built on the latest AWS compute, networking, and disk technologies
  - Amazon FSx can be configured for use with four widely-used file systems

        - Amazon FSx for Windows
            - Fully managed native Windows file system with a wide range of data access, data management, and administrative capabilities
        - Amazon FSx for Lustre
            - Designed for compute intensive applications such as high-performance computing
        - Amazon FSx for NetApp ONTAP
            - Delivers fully managed, multi-protocol shared storage built on NetApp's popular ONTAP file system
        - Amazon FSx for OpenZFS
            - Provides fully managed shared file storage built on the OpenZFS file system
- *AWS Storage Gateway*
    - Hybrid cloud storage service that gives on-premises access to virtually unlimited cloud storage
    - The AWS Storage Gateway service supports three types of gateways
        - File Gateway
            - Provides a seamless way to connect to the cloud in order to store application data files and backup images as durable objects on Amazon S3
        - Volume Gateway
            - Provides block storage volumes that users can mount as iSCSI devices from on-premises application servers

        - Tape Gateway
            - Offers a durable, cost-effective solution to archive users data in the AWS Cloud

- **Backup and Recovery Options**
    - When it comes to backup and recovery options within AWS, there are two main services
    - *AWS Backup*
        - Fully managed backup service that makes it easy to centralize and automate data protection across AWS services
        - The AWS Backup service supports various AWS services
            - Amazon EBS Volumes
            - Amazon RDS Databases
            - Amazon DynamoDB Tables
            - Amazon EFS File Systems
            - AWS Storage Gateway Volumes
        - The user can create backup plans to define the backup requirements and apply these policies to the different AWS resources that it wants to protect
        - AWS Backup service also allows to set backup retention policies that automatically retain and expire backups so that the backup storage costs can be minimized
        - AWS Backup Console is a centralized dashboard to configure the backups

- *AWS Elastic Disaster Recovery*
    - Helps to minimize downtime and data loss with fast, reliable recovery of on-premises and cloud-based applications using affordable, automated, and consistent replications of the system's disks at a low cost
    - This service can initiate the secure data replication on the source servers into a staging area subnet in the user AWS account within the AWS Region selected
    - To recover applications, launch the recovery instances on AWS within minutes using the most up-to-date server state or a previous point in time
- AWS also allows for redundancy across multiple geographical regions and Availability Zones
    - Many AWS services, like Amazon RDS and Amazon EC2, support Multi-Availability Zone deployments
    - *Cross-region Replication*
        - Allows to replicate data across different geographical regions instead of just across availability zones within a single region
- AWS Elastic Disaster Recovery offers reliable recovery for both on-premises and cloud-based applications through consistent replication

- **Retention Lifecycle Policies**
    - The lifecycle of data typically includes four stages
        - *Creation*

        - The process that involves the production and initial storage of the data
    - *Access and Use*
        - Data that is frequently accessed and used for various purposes, including analysis, processing, or transactional purposes
    - *Archival*
        - Process that is undertaken when data is no longer frequently accessed but still needs to be retained for future reference or compliance purposes
    - *Deletion*
        - Process that occurs at the end of the data's lifecycle so that the data can be removed from our systems
- AWS allows users to define lifecycle policies to automate the transition of data between different storage classes and manage their data's eventual deletion
    - The length of data retention will depend on factors such as
        - Business requirements
        - Nature of the data
        - Relevance to ongoing business operations
    - Other considerations
        - Storing data incurs additional costs for the company
- Elastic Block Store services uses snapshots to store data
    - Successive snapshots are incremental and only contain the data that changed since the previous snapshot was created

        - Amazon Data Lifecycle Manager supports these EBS-backed AMIs, which include a snapshot for each EBS volume attached to the source instance
        - Amazon Data Lifecycle Manager applies specific system tags to all snapshots and AMIs created by a policy to distinguish them from snapshots and AMIs created by other methods
    - A lifecycle policy consists of core settings such as
        - Policy type
        - Resource type
        - Target tags
        - Policy schedules
    - By effectively managing the data lifecycle in AWS, businesses can optimize costs, improve operational efficiency, and meet regulatory compliance requirements

- **Configuring S3 Buckets**
    - AWS CloudShell is only available in certain regions
    - Amazon Simple Storage Service (S3)
        - Object storage service that offers a high degree of durability, availability, and scalability

Database Services

Objective 3.4: Identify AWS database services.

- **Databases**
  - *Databases*
    - Organized collections of structured data that are designed to store, manage, and retrieve information
    - The structure of a database determines how data is stored, organized, and manipulated
  - *Structured Databases*
    - Also known as relational databases or SQL databases, are characterized by a highly organized structure
    - Stores data in tables with each table consisting of rows and columns
    - The columns will contain the data attributes and the rows contains the data values for those attributes
    - Some examples of structured, relational, or SQL databases include MySQL, PostgreSQL, and Oracle Database
  - *Semi-Structured Database*
    - Bridges the gap between structured and unstructured databases

        - Some commonly used semi-structured database formats include the XML and JSON data file formats
        - This semi-structured data is known as metadata
    - *Unstructured Database*
        - Also known as non-relational databases or NoSQL databases, can use a variety of data models for accessing and managing data and are designed to handle data without a predefined schema
        - These non-relational databases are optimized for applications that require large data volume, low latency, and flexible data models
        - Non-relational databases can be further classified into key-value databases, document databases, and graph databases
            - *Key-Value Database*
                - Stores data as a collection of key-value pairs, where the key serves as a unique identifier
            - *Document Database*
                - Stores data as JSON objects that are flexible, semi-structured, and hierarchical in nature
            - *Graph Database*
                - Purpose-built to store and navigate relationships

- **Hosted and Managed Databases**

- In the world of databases, there are two methods for running your databases, hosted and managed
- *Hosted Database*
  - Database that is deployed, managed, and maintained on a remote server or cloud infrastructure
  - When you host a database on EC2, your organization is responsible for
    - Provisioning
    - Installation
    - Configuration
    - Maintaining
    - Scaling
    - Backup
    - Recovery
    - Security
- *Managed Database*
  - Database service provided and overseen by a third-party provider or cloud vendor, like AWS
  - The cloud service provider will take on the responsibility for a range of administrative tasks
    - Hardware Provision
    - Software Installation and Updates
    - Backups
    - Scaling

            - Security Measures
            - Routine Maintenance
        - Managed databases are offered with SLAs that guarantee uptime, performance, and other operational service metrics
    - EC2 is most commonly used AWS managed database
        - In EC2 databases, you are in complete control of the installed OS, database version and configuration, and other software components
        - You are also responsible for configuring, securing, maintaining, etc.
    - AWS also provides a lot of different managed database solutions
        - Amazon Relational Database Service (RDS)
            - Fully managed relational database service that provides cost-efficient and resizable capacity while automating time-consuming administration tasks
            - RDS provides you six familiar database engines to choose from
                - MySQL
                - PostgreSQL
                - Oracle
                - SQL Server
                - MariaDB
                - Amazon Aurora
        - *Amazon Aurora*
            - MySQL- and PostgreSQL-compatible relational database service that combines the performance and availability of a traditional

enterprise database with the simplicity and cost-effectiveness of a cloud-based service

    - *Amazon DynamoDB*
        - Fully managed NoSQL database service that provides fast and predictable performance with seamless scalability
    - *Amazon Neptune*
        - Fully managed graph database service that makes it easy to build and run applications that work with highly connected data
    - *Amazon MemoryDB for Redis*
        - Fully managed, Redis-compatible in-memory database service that offers high performance, scalability, and durability

- **Amazon RDS**
    - *Amazon Relationship Database Service (RDS)*
        - Fully managed relational database service provided by AWS
    - RDS supports seven different database engines
        - Amazon Aurora MySQL-Compatible Edition
        - Amazon Aurora PostgreSQL-Compatible Edition
        - MySQL
        - MariaDB
        - PostgreSQL

        - Oracle
        - Microsoft SQL Server
    - Amazon RDS reduces the administrative burden by handling routine database tasks such as provisioning, patching, backup, recovery, failure detection, and repair on behalf of your organization
        - To do this, you can use the AWS Management Console, the Amazon RDS Command Line Interface, or simple API calls to access the capabilities of a production-ready relational database in minutes
        - Amazon RDS allows for safer, simpler, and faster database updates
            - Powering deployments will stay updated with the latest patches
            - It provides best practice guidance and recommendations
    - When setting up your Amazon RDS, you can choose between General Purpose Storage and Provisioned IOPS Storage options
    - Other features
        - Amazon RDS allows to scale the compute and memory resources up to a maximum of 32 virtual CPUs and 244 GB of RAM
        - Amazon RDS provides automated backups, database snapshots, and Multi-AZ deployments for enhanced availability and durability by automatically replacing the compute instances powering your deployment in the event of a hardware failure
        - Data at rest and in transit are encrypted when using Amazon RDS

        - Amazon RDS is considered a pay-per-use model service
        - Amazon RDS is a robust, scalable, and cost-effective solution for managing relational databases in the cloud

- **Amazon Aurora**
    - *Amazon Aurora*
        - Relational database service that offers the speed and availability of high-end commercial databases while maintaining the simplicity and cost-effectiveness of open-source databases
        - Amazon Aurora provides up to 5x the throughput of a MySQL database and 3x the throughput of a PostgreSQL database
    - Amazon Aurora Serverless
        - Auto-scaling configuration that automatically starts up, shuts down, and scales capacity based on the application's needs
    - Amazon Aurora is controlled using
        - the AWS Management Console
        - the Amazon Relational Database Service APIs
    - Major features
        - The storage volumes can be increased in increments of 10 GB up to a maximum of 128 TB

        - Amazon Aurora I/O-Optimized configuration offers cost savings of up to 40%
        - Amazon Aurora allows to create up to 15 database replicas to increase read throughput to support high-volume application requests
        - Amazon Aurora leverages Parallel Query to expedite analytical queries
        - Aurora boasts fault-tolerant storage that is partitioned into 10 GiB chunks and replicated over three Availability Zones
        - Aurora provides versatile backup options
            - Point-in-time recovery
            - User-initiated Database Snapshots in Amazon S3
            - Backtrack feature for Aurora MySQL
    - Aurora supports migrations from multiple database systems
    - Aurora introduces ML and AI functionalities in the database to make real-time predictions possible through SQL
    - When paired with Amazon RDS Proxy, Aurora becomes even more scalable and resilient to enhance your application performance and security
    - Remember, Amazon Aurora is a highly secure, cost-effective, and fully managed relational database service that provides high performance and scalability

- **Amazon DynamoDB**
    - *Amazon DynamoDB*
        - Managed NoSQL database service offered by AWS

- Main features
    - One of its most salient features is its single-digit millisecond performance, meaning it can retrieve or write data in just a few milliseconds
    - DynamoDB scales automatically
    - With DynamoDB, there is no need to worry about server provisioning, patching, or configurations since AWS takes care of these operations in the background
    - As a non-relational NoSQL-database, DynamoDB supports both key-value and document data structures
- Amazon CloudWatch
    - Offers insights into the capacity usage, performance metrics, and operational costs of using the DynamoDB database
- Amazon DynamoDB is a powerful NoSQL database service designed for applications that need scalable, high-performance, and seamless data management

- **Amazon Neptune**
    - *Amazon Neptune*
        - AWS's managed graph database service that is designed specifically for working with highly connected data sets
    - Major features

        - Neptune can support two popular graph models
            - Property Graph
            - Resource Description Framework (RDF)
        - Amazon Neptune is known for its high performance level that is achieved through its purpose-built, high-performance graph database engine
        - Scalability and reliability are also at the forefront in Neptune's design
    - Since Neptune is a managed-database, maintenance and operational tasks are considered to be hassle-free
    - Neptune can be integrated with the AWS Lambda service to create event-driven applications
    - Amazon Neptune is a fully managed graph database service that was designed for applications that need to process and understand intricate and highly connected data sets

- **Amazon MemoryDB for Redis**
    - *Amazon MemoryDB for Redis*
        - AWS's fully managed in-memory database service built on the foundation of Redis
    - *Redis*
        - One of the most popular in-memory key-value data stores used as a database, cache, and message broker service
    - Features

        - One of the main benefits of the Amazon MemoryDB for Redis service is its ability to deliver ultra-fast performance
        - The service provides a fault-tolerant architecture that automatically replicates the data across multiple Availability Zones
        - Amazon MemoryDB for Redis also provides scalability by allowing to scale the clusters up or down based on the application's throughput requirements at any given time
        - AWS will handle tasks like patching, backup, and recovery on behalf of the organization
    - To monitor the database, integrate it with AWS CloudWatch
    - Amazon MemoryDB for Redis offers the perfect blend of performance, resilience, scalability, security, and simplicity

- **Database Migration Tools**
    - Database migrations are an important element in the ever-evolving world of enterprise computing
    - Within the Amazon Web Services cloud, there are two main options for database migrations
        - AWS Database Migration Service (DMS)
            - Designed to move data with minimal downtime for businesses that can't afford long periods of inaccessibility during the migration process

        - This tool supports data sources contained within Oracle, Microsoft SQL Server, MySQL, MariaDB, and other common database types
        - DMS also supports migrations to any of these databases as a target in addition to the database targets provided through Amazon Aurora and other AWS services
    - AWS Schema Conversion Tool (SCT)
        - Complements DMS by facilitating the process of migrating a database schema from one database engine to another
        - SCT assesses the database's schema and the metadata, and then it could provide a report on the actions that are required to make the schema compatible with the target database
        - This tool can automatically convert the majority of the source schema, but there might be custom objects or specific features that might require some manual attention
        - The AWS Schema Conversion Tool can scan on-premise application source code, including Java and .NET
- Both the AWS DMS and the AWS SCT provide the capability to save time and resources when migrating the existing databases to AWS
- The AWS SCT is provided by AWS at no additional cost for the usage and instead, the user only pays for the resources consumed
- The AWS DMS and AWS SCT are powerful and reliable tools when conducting database migrations into the AWS cloud

AI/ML and Analytics Services

Objective 3.7: Identify AWS artificial intelligence and machine learning (AI/ML) services and analytics services.

- **Machine Learning**
  - *Machine Learning*
    - Application or subset of AI that allows machines to learn from data without being explicitly programmed
    - To use machine learning, provide it with a labeled data set where you've already labeled things into categories
    - As we feed it new data, it can label and categorize that data by itself, based upon the patterns it already taught itself
  - To conduct machine learning in the cloud, we have two services in AWS that are focused on machine learning
    - *Amazon SageMaker*
      - Fully managed service that lets data scientists and developers build, train, and deploy machine learning models in the cloud
      - SageMaker offers a suite of tools like
        - Jupyter notebooks for data preprocessing
        - Built-in algorithms for model training
        - One-click deployment functionalities
    - *Amazon Kendra*

- Intelligent search service that is powered by machine learning in the AWS cloud
- Amazon Kendra can also sift through vast data sets, encompassing FAQs, handbooks, technical documents, and many others
- Kendra is really invaluable for large corporations, research institutions, or any organization aiming to streamline its internal search processes for its large knowledge bases and information repositories

- **Image and Video Detection**
    - *Image and video detection services*
        - used to identify objects, faces, text, and other features in images and videos that are fed into the services as input
    - In the AWS Cloud, there are two incredibly powerful services that focus on image and video detection
        - *Amazon Rekognition*
            - Fully managed service that harnesses advanced machine learning models to analyze and detect objects, scenes, and face in images and videos
            - Rekognition can also compare faces, track individuals in videos, and even recognize their emotions or sentiments from their facial features
        - *Amazon Textract*
            - Service that is designed to extract data from various types of documents, powered by machine learning

            - Instead of just reading text from documents, Textract can also be used to understand the content's layout and structure
            - Amazon Textract can operate at scale, and it can handle millions of document pages in just a few hours

- **Language Recognition and Translation**
    - Language processing and translation services play a pivotal role in enhancing communication across the globe by breaking down language barriers and fostering connections
    - In AWS, you will find three services that cater to your language recognition and translation needs
        - *Amazon Polly*
            - Text-to-speech service that uses deep learning to produce lifelike voiceovers from text-based data inputs
            - In a business environment, use Amazon Polly to give a voice to the virtual chat-based assistant on the website
        - *Amazon Transcribe*
            - Utilizes advanced machine learning algorithms to convert audio or video-based speech content into written text
            - Amazon Transcribe can also recognize different speakers within a conversation in audio or video
        - *Amazon Translate*
            - Neural machine translation service that is capable of translating text from one language to another

- **Natural Language Capabilities**

- *Natural Language Processing (NLP)*
    - Method by which computers are trained to understand and interpret human language
- In the AWS Cloud, there are two services that provide natural language capabilities
    - *Amazon Lex*
        - Conversations come to life by enabling the creation of sophisticated, natural language chatbots
        - Amazon Lex provides the tools needed to
            - Builds conversational interfaces
            - Comprehends intent
            - Integrated with other AWS services
        - Fully managed service that helps build conversational interfaces for applications
    - *Amazon Comprehend*
        - Service for diving deep into textual content, which can extract insights that are hidden beneath the surface
        - Amazon Comprehend uses machine learning to uncover insights and relationships within any text provided to the system
        - Natural language processing service that helps extract insights from sets of text-based data

- **Data Analytics and Business Intelligence**

- Data analytics and business intelligence are critically important in modern business strategies since they help us to transform raw data into actionable insights
    - These insights can then be utilized to
        - Make more informed business decisions
        - Drive growth
        - Innovate in alignment
- To cater to these needs, AWS provides us with three powerful services in the realm of data analytics and business intelligence
    - *Amazon Athena*
        - Interactive query service that analyzes vast amounts of data stored in Amazon S3 using standard SQL queries
        - Amazon Athena service is designed to eliminate the need for complex extract-transform-and-load jobs
    - *Amazon Elastic MapReduce (EMR)*
        - Cloud-native big data platform that facilitates processing large amounts of data across resizable clusters of Amazon EC2 instances
    - *Amazon QuickSight*
        - Business intelligence service designed to deliver insights and visualize data
        - Amazon QuickSight has machine learning insight features that can provide machine learning-based business insights automatically
        - QuickSight can be used to gather data from various sources to create a visual dashboard that illustrates user patterns

- **Data Integration and ETL**
  - *Data Integration and Extract, Transform, and Load (ETL)*
    - Enables businesses to extract data from disparate sources, transform into a cohesive format, and load into target databases or data warehouses
  - In the age of data-driven decision-making, having efficient and effective tools for these tasks is crucial, and that is where the AWS Data Exchange service and the AWS Glue service come into play
    - AWS Data Exchange
      - Designed to simplify the process of accessing or sharing data sets in the cloud
    - AWS Glue Service
      - Designed to simplify the process of accessing or sharing data sets in the cloud
      - Glue Data Catalog provides a centralized metadata repository to make ETL jobs faster and more accessible

- **Data Streaming and Search**
  - Data Streaming
    - Real-time processing of high volumes of fast-moving data that enables businesses to gather actionable insights
    - Search services help businesses sift through vast volumes of data to find exactly what they're looking for in a fraction of a second

- Inside of AWS, there are four main services that are used for data streaming and search
  - *Amazon Kinesis*
    - AWS's flagship data streaming service that allows users to easily collect, process, and analyze real-time streaming data to derive insights and respond quickly to new information
  - *Amazon Managed Streaming for Apache Kafka*
    - Fully managed service that helps users set up, scale, and operate Apache Kafka clusters in AWS
    - *Apache Kafka*
      - Open-source stream-processing software platform that was originally developed by LinkedIn and then donated to the Apache Software Foundation
  - *Amazon OpenSearch*
    - Managed service that makes it easy to deploy, operate, and scale OpenSearch for log analytics, application monitoring, and other relevant use case
  - *Amazon Redshift*
    - Fully managed, petabyte-scale data warehousing service that offers lightning-fast queries using SQL, extract-transform-and-load, and Business Intelligence tools
    - Redshift is also designed to integrate seamlessly with various data loading and business intelligence tools

Security Capabilities

Objective 2.4: Identify components and resources for security.

- **Security Groups**
    - *Security Groups*
        - Function as virtual firewalls designed specifically to protect EC2 instances
    - Two basics types of firewalls
        - *Host-based Firewall*
            - Installed on individual servers or devices and controls incoming and outgoing network traffic specific to that machine, based on a set of configured rules
        - *Network-based Firewall*
            - Positioned at a strategic point within a network, usually at the boundary between internal and external networks, and it controls traffic entering or leaving the entire network
    - In the AWS cloud, security groups act as a host-based firewall for EC2 instances
    - Security groups function on the principle of rules
        - These rules dictate which traffic can enter or leave a particular EC2 instance
        - Two Types of Rules
            - Inbound Rules
                - Dictate which incoming traffic is allowed into a given instance

            - Don't always block all traffic that is destined for an EC2 instance
        - Outbound Rules
            - Used to regulate the outgoing traffic from an EC2 instance
    - To protect these EC2 instances from any malicious Internet-based threats, we could define Security Group rules that only allow HTTP and HTTPS traffic to be received by the EC2 instances and all other types of traffic would be blocked by default
        - Set Security Group rules to only allow traffic from specific EC2 instances or IP addresses to that managed database
        - Security groups cannot filter traffic based on domain names, it can only allow or deny traffic based on ports and IP addresses
        - Security groups are essentially a virtual firewall that can be used with EC2 instances

- **Network ACLs**
    - *Network Access Control List (ACL)*
        - Offers a layer of security that operates at the sub-net level to grant users the flexibility to determine the traffic flow between various subnets
        - Network ACLs function as stateless packet filters
    - Within Amazon VPCs, configure network ACL through the creation of inbound and outbound rules for each of the subnets in the VPC

- Best practices
    - Always write the network ACL rules from the most specific to the most broad
    - Understand the default configuration that everything will be set to
    - Network ACLs are created in Amazon VPCs to act as a network firewall and serve as a gatekeeper to manage traffic at the subnet level into or out of the VPC

- **AWS WAF**
    - *AWS Web Application Firewall (AWS WAF)*
        - Security solution that monitors and filters incoming web traffic to protect web applications from potential threats and malicious attacks
        - The AWS WAF is a virtualized web application firewall service that monitors the HTTP and HTTPS traffic going to and from a given web application
    - AWS WAF is designed to shield applications against some of the most widespread web exploits seen on the Internet
        - It is not considered to be a one-size-fits-all solution
        - It can also be combined with monitoring capabilities using Amazon CloudWatch

        - It is beneficial to protect web applications, regardless of the industry
    - AWS WAF is a dynamic security solution that scrutinizes HTTP and HTTPS traffic and allows users to define conditions to block, allow, or monitor web requests

- **Encryption Options**
    - *Encryption*
        - The process of converting information into a code to prevent unauthorized access or viewing
    - Encryption is used with three different data states
        - *Data at Rest*
            - Data that is stored on physical media, such as hard drives or databases, and is not actively being transmitted
            - AWS has a wide variety of services that can help us to enforce data at rest encryption, including the AWS Key Management Service known as KMS
                - Key Management Service (KMS) helps facilitate the centralized management of cryptographic keys used for data at rest encryption requirements
        - *Data in Transit*
            - Data that is actively moving between two or more systems, such as over the Internet or through a private network
            - Use encryption to create a secure pathway or virtual tunnel to send data from one system to another over an otherwise untrusted network like the Internet

            - To protect our data in transit, many of the AWS services like AWS Elastic Load Balancing, Amazon CloudFront, Amazon S3, and Amazon RDS have built-in capabilities to enable encryption of our data in transit by default
        - *Data in Processing*
            - Data currently being used or processed by applications, and usually being held in systems' RAM or processor caches
            - By maintaining a secure environment, regular patching, and following a principle of least privilege, data remains secure even during processing

- **AWS Trusted Advisor**
    - *AWS Trusted Advisor*
        - Vital tool that offers guidance to AWS users to optimize their cloud-based infrastructure
    - It is like your own personalized cloud consultant that will continuously monitoring your AWS resources for you and provide you with recommendations on the areas that matter most to your organization
        - Cost Optimization
            - This service can identify any idle and underutilized resources and propose the various ways to reduce costs without compromising the efficiency of the services
        - Security

            - The service can check for unrestricted ports, weak passwords, and other potential security threats and then it will offer some recommendations to improve the security of the cloud environment
        - Performance
            - Ensures the services are running optimally by analyzing usage patterns and configurations, and then suggesting enhancements to improve the system's performance
        - Fault Tolerance
            - The Trusted Advisor can evaluate the AWS environment for resilience against failures and then make recommendations to improve the fault tolerance
        - Service Limits
            - The AWS Trusted Advisor can keep an eye on these limits, and then it can alert anytime the user is getting close to hitting one of those limits
    - The AWS Trusted Advisor is a personalized cloud consultant constantly monitoring the AWS resources and also helps with fine-tuning those resources

- **AWS Security Information**
    - *AWS Knowledge Center*
        - Curated by experts who work at AWS, and serves as a repository filled with articles, white papers, best practices, and technical documentation

        - The AWS Knowledge Center also deepens the understanding of the different AWS services and specific features
    - *AWS Security Center*
        - Full of information, tools, and resources dedicated to the security best practices that are used in AWS
    - *AWS Security Blog*
        - Dynamic platform to find insights shared by AWS security experts
        - The AWS Security Blog provides a fresh and continuous stream of the best security information
    - Remember, when navigating the vast world of AWS security, there are three main sources of information you should consider
        - AWS Knowledge Center
            - Has all of the detailed documentation and solutions you are looking for
        - AWS Security Center
            - Hosts the best practices and security tools you may need to use
        - AWS Security Blog
            - Has insights and updates from the frontline of AWS security

- **Third-party Security Products**
    - *Third-party Security Products*

    - Security solutions that are provided by vendors outside of AWS
- *AWS Marketplace*
    - Curated digital catalog that allows AWS customers to find, test, purchase, and deploy software from external providers
- *Threat Detection*
    - Focuses on detecting threats in real-time by continuously monitoring network traffic, system logs, and application activity, and flagging anomalies or suspicious activities for review
- *Data Protection*
    - Software, applications, or systems designed to safeguard data from loss, corruption, theft, unauthorized access, or other potential threats
    - Data protection is also focused on safeguarding data at rest, in transit, and in processing
- *Access Control*
    - Process of determining who is allowed to access specific resources and what actions they can perform
- *Compliance Management*
    - Process of ensuring an organization's actions and procedures adhere to external laws, regulations, and industry standards, as well as internal policies and best practices

- Depending on your industry, your organization may be required to operate under stringent regulatory frameworks like HIPPA, Sarbanes-Oxley, FISMA, or other relevant laws and regulations

Governance and Compliance

Objective 2.2 Understand AWS Cloud security, governance, and compliance concepts.

- **Compliance in AWS**
  - *Compliance*
    - The act of adhering to established guidelines, standards, or laws that have been set forth by regulatory bodies or authorities
    - Compliance requirements are non-negotiable when dealing with critical and sensitive data within the organization
  - In AWS, there are many different tools and services that you can use to ensure you are meeting your compliance requirements
    - *AWS Compliance*
      - Used to ensure that the cloud platform aligns with the best practices, standards, and regulations that apply in the organization's industry
      - AWS services and resources remain compliant with various global standards and regulations
        - PCI-DSS
        - HIPAA
        - HITECH
        - GDPR
      - The AWS Compliance tool is designed to bolster the trust and confidence that the organization and others place in AWS to manage and store their critical data

        - AWS compliance supports 143 different security standards and compliance certifications
        - The organization can inherit many of the security controls that have been put in place by AWS in order to reduce the overhead and decrease the deployment times
    - *AWS Artifact*
        - Designed to be the one-stop destination for accessing any compliance reports and certifications maintained by AWS
  - Cloud compliance ensures the organization has comprehensive security program that meets all of its legal, regulatory, and internal requirements

- **Monitoring in AWS**
  - *Monitoring*
    - The systematic process of observing, collecting, and analyzing data to track and assess the performance and health of systems or applications
    - Monitoring is critical for maintaining the health, availability, and performance of the organization's cloud resources
  - *Amazon CloudWatch*
    - Allows users to collect and track metrics, set up alarms, and monitor the health and performance of their AWS resources and applications in real time
    - CloudWatch can gather logs from the EC2 instances, Lambda functions, or even on-premise servers

        - CloudWatch has storage and retrieval capabilities so that the user can look at the data and metrics collected in the past and analyze those trends and patterns over time
    - Monitoring is a process that involves observation, data collection, and analysis

- **Auditing in AWS**
    - *Auditing*
        - The systematic process of evaluating and verifying an organization's operations or systems to ensure their accuracy and compliance with regulations, standards, and best practices
    - In AWS, we have three main tools that we use to conduct auditing in the AWS Cloud
        - *AWS CloudTrail*
            - Service used to capture and log all API calls made within an AWS account to provide a detailed history of all the account activity for later security analysis and compliance auditing
            - With CloudTrail, you can get the details of the API calls made on your account
                - Source IP
                - Event
                - Timestamps
        - *AWS Audit Manager*

            - Managed service that is focused on streamlining the process of assessing how well the environment aligns with specific regulations, industry standards, and best practices
        - *AWS Config*
            - Continuously tracks and monitors any changes to the AWS resources
    - Auditing is critically important when the organization operates in a regulated industry or location

- **AWS Security Hub**
    - *AWS Security Hub*
        - Essential service within AWS that is dedicated to providing a unified view of the security and compliance landscape
        - AWS Security Hub can also be used with other integrated, third-party partner solutions within AWS
        - The primary purpose of the AWS Security Hub is to provide a centralized security command center
    - *"Single Pane of Glass" Solution*
        - Management console or dashboard that integrates information from multiple sources into one accessible and user-friendly interface
    - The AWS Security Hub continuously monitors the environment using automated security checks based on the AWS best practices and industry standards

        - Once it identifies a potential issue or vulnerability, the hub will provide
            - Detailed findings
            - Current status
            - Recommendations
        - AWS Security Hub allows for the integration of third-party services that are supported by different AWS partners
    - The AWS Security Hub is a central dashboard that consolidates security findings across AWS services and integrated partner solutions

- **AWS Inspector**
    - *AWS Inspector*
        - Helps with analyzing applications, detecting potential vulnerabilities, and offering comprehensive reports to prioritize and remediate the detected vulnerabilities
        - This service scrutinizes the running applications hosted by AWS and then provides a detailed assessment report with all of the potential security issues
    - How does the AWS Inspector service work?
        - The service performs an in-depth assessment of your application's behavior as compared to the industry's known best practices and security standards
        - A detailed report is generated by the AWS Inspector service

- Use the AWS Inspector service to automatically run scans of the cloud-hosted applications periodically
- The AWS Inspector is a powerful tool in the AWS security toolkit

- **AWS GuardDuty**
  - *AWS GuardDuty*
    - Threat detection service that provides continuous monitoring capability which is paired with machine learning and threat intelligence to detect and alert on suspicious behavior that might indicate a potential security threat
  - AWS GuardDuty relies on three primary data sources
    - AWS CloudTrail Event Logs
    - Amazon VPC Flow Logs
    - Route53 DNS Logs
  - Consider using AWS GuardDuty as part of the organization's security architecture to secure AWS cloud resources

- **AWS Shield**
  - *AWS Shield*
    - Managed service that provides protection against Distributed Denial of Service attacks for applications running in the AWS cloud
  - *Distributed Denial of Service (DDoS) Attack*

    - Malicious attempt to disrupt the normal functioning of a network, service, or website by overwhelming it with a flood of Internet traffic from multiple sources
    - Use a DDoS protection service to divert the malicious incoming traffic to a specialized environment that can deal with it
  - AWS Shield is a managed DDoS protection service that is engineered to keep the applications resilient against any incoming DDoS attacks
    - AWS Shield can provide detailed insights into ongoing attacks
    - AWS redirects that traffic so it won't negatively affect the operations
  - AWS Shield service provides two levels of protection
    - *AWS Shield Standard*
      - Free service that is available to all AWS customers, and it provides protection against most common attacks
      - Medium-sized business with limited budget
    - *AWS Shield Advanced*
      - Provides enhanced DDoS protection, 24/7 access to the AWS DDoS response team, and additional cost protection
      - Higher security needs, running mission-critical application or at high risk
  - AWS Shield is able to prevent very large-scale DDoS attacks on our behalf
  - According to AWS, they prevent around 200 to 250 DDoS attacks every day
  - The decision of whether to use AWS Shield Standard or AWS Shield Advanced depends on the specific needs and budget

Other AWS Services

Objective 3.8: Identify services from other in-scope AWS service categories.

- **Application Integration Services**
    - *Application Integration Services*
        - Ensure seamless communication and coordination between applications and microservices occur
    - There are four main application integration services in AWS
        - *Amazon EventBridge*
            - Serverless event bus that facilitates the connection of application by using data from various sources
        - *Amazon Simple Notification Service (SNS)*
            - Robust, fully managed publish and subscribe messaging service
        - *Amazon Simple Queue Service (SQS)*
            - Ensures the proper communication between decoupled components of cloud applications occurs
        - *Amazon Step Functions*
            - Allow the coordination of multiple AWS services into serverless workflows

- **Business Productivity Services**
    - *Business Productivity Services*
        - Act as the backbone for companies that want to boost efficiency and optimize workflows

- AWS offers two reliable solutions to help increase your business' productivity
    - *Amazon Connect*
        - Designed to replace traditional call center infrastructure with a cloud-based infrastructure that offers an intuitive, scalable solution for setting up and managing customer service call centers
        - Amazon Connect provides integration capabilities to tailor the service to provide a personalized experience for each caller or chat participant
        - Manages call volumes and creates chatbots
            - *Amazon Transcribe*
                - Automatic transcriptions of voice calls
            - *Amazon Translate*
                - Automatic translation of chat data
            - *Amazon Polly*
                - Responds to customers using a virtualized voice attendant
    - *Amazon Simple Email Service (SES)*
        - Cloud-based email sending service designed to help digital marketers and application developers send marketing, notification, and transactional emails
- Amazon Connect provides us with an all-in-one solution for creating a modern customer service call center that includes both voice and chat capabilities

  - Amazon SES is focused on ensuring that your emails are delivered effectively while maintaining the integrity and reputation of your domain when sending those emails

- **Customer Engagement Services**
  - *Customer Engagement Services*
    - Instrumental in helping to nurture the relationship between AWS and the organization by providing enhanced support
  - Within AWS, there are three services that are focused on customer engagement
    - *Activate for Startups*
      - Designed to help new companies begin to use AWS as they get ready to scale
    - *AWS IQ*
      - Marketplace that connects AWS customers with AWS Certified experts for on-demand project work that organizations might need to complete
    - *AWS Support*
      - Comprehensive support program used to provide 24/7 technical assistance to AWS customers

- **Developer Tools and Services**
  - *Developer Tools and Services*
    - Provide developers the tools and services they need to architect, code, test, and deploy applications in a more streamlined and efficient manner

- AWS has a lot of different developer tools and services to support the entire software development lifecycle
    - *AWS AppConfig*
        - Tool that lets developers and IT administrators safely deploy application configurations in real time
    - *AWS Cloud9*
        - Integrated Development Environment (IDE) that operates in the cloud
    - *AWS CloudShell*
        - Service that grants developers command-line access to AWS directly from within user's AWS Management Console
    - *AWS CodeArtifact*
        - Managed artifact repository service that lets teams store, publish, and share software packages used in their development process
    - *AWS CodeBuild*
        - Fully managed build service that allows developers to compile, test, and deploy code without the need to provision or manage servers
    - *AWS CodeCommit*
        - Secure, scalable, and managed source control service that can be used to host private Git repositories
    - *AWS CodeDeploy*
        - Deployment service that automates application deployments into various compute services such as EC2 and Lambda
    - *AWS CodePipeline*

        - Continuous integration and continuous delivery service, known as a CI/CD service, that can be used to improve the organization's release processes
    - *AWS CodeStar*
        - All-encompassing platform that is tailored for the creation, building, and deployment of applications within the AWS environment
        - AWS CodeStar is also used to create a complete CI/CD pipeline
    - *AWS X-Ray*
        - Service that offers insights into the behavior of applications and makes it easier to analyze and debug applications

- **Frontend Web and Mobile Services**
  - *Frontend development*
    - Process of designing and creating the visual and interactive elements of a website or mobile application
      - Layout
      - Design
      - Graphics
      - User Interface Components (Buttons, Menus, and Forms)
  - While backend development focuses on databases, servers, and application logic, frontend development prioritizes user experience to ensure that the application is responsive, intuitive, and visually appealing across various devices and screen sizes
  - In AWS, we have three main frontend web and mobile services

        - *AWS Amplify*
            - Acts as a bridge between frontend web and mobile developers and the powerful tools that AWS provides
            - At its core, AWS Amplify, is a development platform that provides a robust set of tools
                - CLI
                - Authentication Features
                - API Access
                - Storage Capabilities
                - AI/ML
            - Amplify provides ready-to-use libraries that can reduce development time and ensure apps are functional, scalable, and secure
        - *AWS AppSync*
            - Managed GraphQL service designed to provide real-time data
        - *AWS Device Farm*
            - Fully managed testing service that allows developers to test mobile applications across a vast array of real devices
            - With AWS Device Farm, users can test and ensure their application’s compatibility across numerous devices and across multiple operating systems

- **Internet of Things (IoT) Services**
    - *Internet of Things (IoT)*
        - Growing network of connected devices that collect and share data

        - IoT refers to the network of physical objects embedded with sensors, software, and other technologies that connect and exchange data over the Internet
    - In AWS, there are two main services that you should be aware of when developing applications that will interact with the Internet of Things
        - *AWS IoT Core*
            - Managed cloud service that lets connected devices easily and securely interact with cloud applications and other devices
            - AWS IoT Core enables the seamless communication between devices and the AWS Cloud
        - *AWS IoT Greengrass*
            - Allows devices to perform local data processing and run AWS Lambda functions offline
    - Remember that the AWS IoT Core ensures your devices are always talking, and the AWS IoT Greengrass is used to ensure that your devices are never truly offline

- **Management and Governance Services**
    - *Management and Governance Services*
        - Provide tools to oversee, administer, and optimize AWS resources to ensure organization's compliance with organizational policies and industry best practices
    - Within AWS, there's a wide variety of tools that are devoted to management and governance

- *Auto Scaling*
  - Ensures that applications remain up and running by automatically adjusting their capacity based on the conditions defined
  - Auto scaling can be applied both vertically and horizontally
    - *Vertical Scaling*
      - Scaling up or scaling down
    - *Horizontal Scaling*
      - Scaling out or scaling in
- *CloudFormation*
  - Allows users to define and provision AWS infrastructure resources using templates written in a JSON or YAML file
  - CloudFormation can create massive and complex systems
- *Compute Optimizer*
  - Offers recommendations for resources to ensure they are optimized for performance and cost-effectiveness
- *Config*
  - Focused on tracking resource inventory and changes to ensure compliance and security in a cloud environment
- *Control Tower*
  - Designed to automate the setup of a well-architected multi-account AWS environment
- *Health Dashboard*
  - Provides real-time information on the operational status and performance of AWS services and infrastructure across all regions

        - *Launch Wizard*
            - Simplifies the deployment of applications on AWS using guided, best-practice-driven configurations
        - *License Manager*
            - Streamlines the management and governance of software licenses across AWS and on-premise environments
        - *Resource Groups*
            - Allows users to organize AWS resources based on criteria and operational needs
        - *Tag Editor*
            - Enables the bulk addition, modification, or deletion of resource tags across AWS services
        - *Service Catalog*
            - Allows organizations to create and manage approved catalogs of resources that are available for use on AWS

- **Security, Identity, and Compliance Services**
    - Security, Identity, and Compliance Services provides the tools and features needed to help manage access, protect data, and ensure consistent compliance across cloud-based environments
    - AWS provides a number of different security, identity, and compliance services
        - *Audit Manager*
            - Simplifies the assessment of the AWS workload compliance by continually auditing AWS usage to ensure that it aligns with specific regulatory standards
        - *AWS Certificate Manager*

        - Eliminates the complexities of managing TLS and SSL certificates by handling the certificate provisioning, deployment, and renewal on behalf of the organization
    - *CloudHSM*
        - Provides the capability of having our own dedicated Hardware Security Module in the cloud
            - Hardware Security Module
                - Physical device that is designed to securely manage, generate, and store cryptographic keys for the organization
    - *AWS Cognito*
        - Service that provides secure user sign-up, sign-in, and access control to the web and mobile applications
    - *AWS Detective*
        - Service that helps analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities within the AWS accounts
    - *AWS Directory Service*
        - Used to provide a managed Microsoft Active Directory and other directory capabilities
    - *AWS Firewall Manager*
        - Allows for the centralized configuration and management of AWS Web Application Firewall rules, AWS Shield Advanced protections, and VPC security groups across multiple accounts and applications
    - *AWS Key Management Service*

  - Provides a centralized control for the cryptographic keys
- *AWS Macie*
  - Uses machine learning to automatically discover, classify, and protect sensitive data within AWS
- *AWS Network Firewall*
  - Managed network security service that provides granular control over inbound and outbound traffic to help protect virtual networks and applications
- *AWS Resource Access Manager*
  - Allows to share AWS resources securely across multiple AWS accounts and organize resource access for collaboration and resource optimization

Technical Resources and Support

Objective 4.3: Identify AWS technical resources and AWS Support options.

- **AWS Documentation**
    - *AWS Documentation*
        - Comprehensive resource that can provide detailed information, guides, and tutorials for using the different cloud computing services and solutions in AWS
    - Within AWS, there are three primary types of documentation that you will use
        - *White Papers*
            - Carefully created documents that are designed to dive deeply into the technical nature of the various AWS services and architectures
            - These white papers are created by AWS experts who ensure that users are receiving the most accurate, up-to-date, and actionable insights possible for the different AWS services
            - The user can also sometimes find a single white paper for a given architecture or solution
        - *Blogs*
            - Go-to source for staying up-to-date with the ever-evolving world of AWS and its services
            - Regularly check the AWS Blogs to gain insights into new features, tools, or services that might be useful in helping to achieve the organization's goals
        - *Service Documentation*

        - For each AWS service, there is a dedicated documentation section that contains detailed information, tutorials, and API references
    - Many people believe that the AWS White Papers and Service Documentation are the same thing, but they are not
        - The AWS white papers are in-depth technical documents that explore specific topics, concepts, best practices, and architectural patterns related to cloud computing and AWS services
        - These white papers are generally more conceptual and strategic in nature than the AWS service documentation
        - The AWS service documentation, on the other hand, provides detailed, practical information about individual AWS services
        - This service documentation is more hands-on and directly applicable to a specific task
    - AWS Documentation
        - Understanding, implementing, or staying up-to-date
    - AWS White Papers
        - Technical documents with in-depth insights, best practices, and architectural guidance

- **AWS Technical Resources**
    - *AWS Technical Resources*

        - Detailed guides, insightful articles, and real-world experiences that have been shared by AWS enthusiasts to help our users navigate the large ecosystem of AWS services and tools more effectively
    - Within the AWS cloud, there are three technical resources that we commonly utilize
        - *AWS Prescriptive Guidance*
            - Designed to provide the best practices, recommended configurations, and in-depth instructions needed for users to best leverage their AWS services
        - *AWS Knowledge Center*
            - Vast library filled with articles, white papers, best practices, and technical documentation that has been created and maintained by the experts at AWS
        - *AWS re:Post*
            - Dedicated blogging platform that AWS enthusiasts and AWS employees can both use to share their insights, updates, and experiences related to the various AWS services

- **AWS Support Options**
    - AWS Support Options
        - Helps to ensure that AWS users can receive the help and assistance they need at any point during their cloud journey
    - There are numerous different AWS Support Options that you should be aware of
        - *Customer Service and Communities*

            - Through this platform, AWS provides guidance on any of the account and billing questions, as well as providing the ability to request service limit increases
        - *AWS Developer Support*
            - Tailored for those who are developing or testing their applications on AWS
        - *AWS Business Support*
            - Premium support tier that offers fast response times, infrastructure event management, 24/7 phone and email access to cloud support engineers, and many other features
        - *AWS Enterprise On-Ramp Support*
            - Designed for organizations that are newer to the cloud and are looking to gain confidence and experience when using AWS
        - *AWS Enterprise Support*
            - Offers all the benefits of the Business Support tier, plus it provides others like Infrastructure Event Management, a dedicated Technical Account Manager, and periodic infrastructure health checks
    - AWS Support Options are not just about troubleshooting and resolving issues, but instead they are really focused on empowering organizations users when operating in the AWS cloud

- **Cost Optimization**
    - *Cost Optimization*

        - Involves strategically managing and controlling the AWS resources and services to reduce expenditure while maintaining or enhancing performance and efficiency
    - To help with this, AWS has three main services that can be used to ensure you aren't wasting money on resources that you don't need in your cloud-based environment
        - *Trusted Advisor*
            - Operates in real time and can offer some valuable insights for optimizing AWS resources, increasing security measures, and achieving better performance and reliability
            - Trusted Advisor service can point out any areas with wasted resources and then it can suggest optimization measures
        - *AWS Health Dashboard*
            - Serves as a holistic view of the operational health of AWS services being leveraged in the cloud-based environment
        - *AWS Health API*
            - Application programming interface that acts like a conduit for users to be able to programmatically extract data from the AWS Health Dashboard

- **AWS Trust and Safety**
    - AWS has dedicated a lot of resources through the AWS Trust and Safety team to ensure the safe and compliant use of its services
    - *AWS Trust and Safety Team*

        - Responsible for protecting AWS customers, partners, and Internet users from the bad folks who attempt to use AWS services for abusive or illegal purposes
        - The AWS Trust and Safety team mission is to detect and prevent misuse or abuse of AWS services
    - *AWS Acceptable Use Policy*
        - Comprehensive document that delineates the do's and don'ts for users in the AWS cloud to ensure that AWS remains a secure, efficient, and user-friendly environment
    - The team can ensure that the robustness and integrity of the AWS infrastructure are safeguarded from most potential threats
    - The AWS Trust and Safety team does not operate in a silo, but instead, they actively collaborate with the customer organization
        - The Trust and Safety team will look into the matter and provide guidance on how to ensure that the organization's data remains safe, secure, and confidential
    - The AWS Trust and Safety team is responsible for protecting AWS customers, partners, and Internet users from threat actors

- **AWS Partner Network**
    - *AWS Partner Network (APN)*

        - Global community of businesses that provide a variety of services, software, and resources that complement AWS offerings to help customers more effectively use AWS services
    - The AWS Partner Network really focuses on the AWS Marketplace, independent software vendors, and AWS system integrators
        - *AWS Marketplace*
            - Digital catalog that connects AWS customers with third-party software, data, and services that are compatible with the AWS platform
            - One such AWS Marketplace solution you could choose is the cloud-native endpoint protection platform known as CrowdStrike Falcon
                - CrowdStrike Falcon
                    - Software that can integrate seamlessly with other AWS services to provide enhanced visibility and protection against sophisticated threats, malware, and vulnerabilities
        - *Independent Software Vendors*
            - The creators of unique software that is tailored for use within the AWS cloud
        - *AWS System Integrators*
            - The experts who can architect, design, and connect different services together so that they communicate smoothly, seamlessly, and efficiently

- Being an AWS Partner can have a lot of other benefits
  - Partner Training and Certification
    - AWS offers exclusive training sessions for its partners to ensure that they are well-versed with the latest information on the various AWS services available for use
  - Partner Events
    - AWS Partner Events offer a place where partners can network with each other, share knowledge, and collaborate on interesting new projects and solutions
  - Partner Volume Discounts
    - If the organization uses a lot of AWS services, loyalty can be rewarded with some Partner Volume Discounts
  - Recognition and Credibility
    - By gaining the certification, the user is instantly recognizable as someone who has a certain level of knowledge about the various AWS services and this also leads to increased credibility

- **AWS Marketplace**
  - *AWS Marketplace*
    - Serves as a solution hub where the user can find a wide variety of third-party software and services that perfectly align with AWS's

infrastructure to create a seamless and efficient cloud computing environment

- The AWS Marketplace is a great tool that provides us with a lot of benefits
  - Streamlined Software Discovery and Deployment
    - Customers can find a large catalog of software products that span categories such as data analytics, security, networking, and more
  - Robust Cost Management Solutions
    - There are vendors who offer their products with many different pricing models, including a pay-as-you-go model, a subscription-based model, or simply a free trial-based model
  - Enhanced Governance and Entitlement Options
    - With entitlement management, the organization can monitor and manage the usage of software products acquired from the marketplace to ensure compliance with licensing agreements
- Private Marketplace
  - Used by larger, enterprise-level organizations to create a customized

    digital catalog to manage and govern their software procurement
- The AWS Marketplace serves as a dynamic platform that caters to both AWS customers and third-party vendors to foster a vibrant ecosystem where software products can be discovered, purchased, and managed

- **AWS Support Center**
  - *AWS Support Center*

        - Centralized hub where AWS users can access a range of resources, including technical support, expertise, and tools to help manage and optimize their AWS services and applications
    - 4 Most Popular Support Services
        - *Support Plans*
            - Tiered customer service offerings provided by AWS that contain a wide range of services from basic support and access to AWS documentation to 24 hours per day access to Cloud Support Engineers, infrastructure event management, and technical account management
            - Because AWS understands that each organization's needs are unique, it offers a range of support plans tailored for different use cases
        - *AWS Support Engineers*
            - Experts in cloud computing who are highly skilled and ready to assist with any technical issues
        - *AWS Professional Services & AWS Solutions Architects*
            - They will help to create, design, implement the cloud project to achieve success, and work closely with organizations to implement the architectural best practices
        - *AWS Support Forum*
            - Community-driven platform where AWS users can ask questions, share knowledge, and collaborate on solutions regarding various AWS services and technologies

Conclusion

- **Conclusion**
    - 4 Domains of AWS Certified Cloud Practitioner
        - Domain 1: Cloud Concepts
            - It makes up 24% of the exam
        - Domain 2: Security and Compliance
            - It makes up 30% of the exam
        - Domain 3: Cloud Technology and Services
            - It makes up 34% of the exam
        - Domain 4: Billing, Pricing, and Support
            - It makes up 12% of the exam
    - How do you sign up and schedule your exam?
        - Pearson VUE
            - You can take it at any Pearson VUE testing center worldwide, at either a local testing center or online
            - You can buy that exam voucher by going to Pearson Vue directly when you're scheduling your exam at **pearsonvue.com**, or going to the voucher store at lpi.org to buy it from their online store
            - Pearson VUE and LPI have now created a capability for you to take your certification exam online from the comfort of your home or office, using the Pearson VUE OnVue testing system

        - AWS
            - Go directly to the AWS certification page at **aws.training/certification**
    - Amazon does NOT provide you with your pass or fail results immediately
        - It usually takes anywhere from 24 to 48 hours
    - Top five tips for increasing your score on the exam
        - Use a cheat sheet
            - You're not allowed to actually carry anything into the exam with you, but if you're at a local testing center, they will give you a whiteboard or a dry erase sheet that's about the size of a normal piece of paper
            - Once the clock starts on the exam, you can brain-dump anything you want onto that paper
            - Use the sheet and spend the first 1-2 minutes writing down those important things you may forget later on
        - Skip any questions that are giving you trouble
            - If you find yourself struggling with a really hard question, just mark it for review and skip it
            - Students who do this end up increasing their score by at least 5% to 10% over their peers who try to do the simulations at the beginning of their exam
        - Take a guess

            - If you're in doubt, I want you to take a guess from the possible answer choices
            - There is no penalty for guessing incorrectly on the exam
            - If you are in doubt of the right answer, try to eliminate as many choices as possible and guess between the remaining answer options
        - Pick the best time for your exam
            - Pick the time of day that works best for you
            - Don't try to squeeze the exam in after working a long day at the office
        - Be confident
            - You've got this!
            - You should already know you're going to pass!
            - You should have already studied all the information in this course, you've watched the videos, you've taken the quizzes, you've studied your downloadable study notes
            - If you're not confident right now, then wait a few days to schedule your exam
            - Take a bunch of practice exams and build up your confidence
    - When you take a practice exam, your goal is not to memorize the answer key
        - You need to understand why the right answer was right and the wrong answers are wrong

- Good luck, and I hope to see you again in a future course as you continue upwards in your cloud computing career and continue to climb up the certification ladder!