Module 1: Introduction

Cloud Computing

Client-server model
- all basic apps operate on a client-server model
- customer makes a request
- with permissions, the server responds to the request
- client -> web browser or app
- server -> EC2 (Amazon Elastic Compute Cloud)
"You only pay for what you use"
- capacity constraints
- instances: adjust dependencies based on requirements

Types of deployment
- Cloud-based deployment
  - existing applications: can be ran and deployed to the cloud
  - new applications: can be designed and built in the cloud
- On-premises deployment
  - aka private deployment
  - deploy resources by using virtualization and resource management tools
  - increase resource utilization by using app management and virtualization technologies
- Hybrid Deployment
  - connect cloud-based resources to on-premises infrastructure
  - integrate cloud-based resources with legacy IT applications

Model 2: Compute in the Cloud

- AWS has already: built datacenters, secured datacenters, purchased servers, installed servers
- only pay for running instances
- sharing the host and resources with multiple other instances aka virtual machine
- Multitenancy: sharing underlying hardware between virtual machines
- configurations
  - Windows, Linux
  - Internal business apps, web apps
  - databases
  - third-party software
- Networking

○ private or public request types

Amazon EC2 Instance Types

General purpose
- Balanced resources
- Diverse workloads - web servers, code repositories etc.

Compute optimized
- offers high-performance processors
- for intensive tasks - gaming servers, high performance computing (HPC), scientific modeling etc.

Memory optimized
- memory intensive tasks
- high performance databases etc.

Accelerated computing
- using hardware accelerators
- floating point number calculations, graphics processing, data pattern matching etc.

Storage optimized
- high performance for locally stored data

Amazon EC2 Pricing

On-Demand
Reserved Instances
EC2 Instance Savings Plans
Spot Instances
Dedicated Hosts

	Description	Use Cases
On-Demand	instances run continuously until you stop them, and only pay for compute time you use	short-term, **irregular** workloads that **cannot be interrupted**
Reserved Instances	two types: **Standard** and **Convertible** billing discount applied to the use of On-Demand instances in your account	when you **know** the instance type and size, platform description (operating system), tenancy
EC2 Instance Savings Plans	reduce EC2 instance costs when you make an hourly spend commitment to an instance family and REgion for a 1-year or 3-year term	workloads that require **flexibility** in Amazon EC2 **usage** over the duration of the commitment term
Spot Instances	use unused Amazon EC2 computing capacity (offers savings up to 90% off of	workloads with **flexible** start and end times, that can **withstand**

	On-Demand prices)	**interruptions**
Dedicated Hosts	physical servers with instance capacity fully dedicated to your use	big budget workloads (most **expensive**!!)

Scaling Amazon EC2

Amazon EC2 Auto Scaling
- Scalability - start out with only necessary resources and architecture is respondent to changing demand
- Auto scaling enables automatic adding or removing instances in response to changing application demand
    - instances are scaled in and out as needed automatically
    - Two types:
        - Dynamic scaling
        - Predictive scaling
    - Dynamic and predictive scaling can be used together for max efficiency
- Minimum capacity
    - minimum number of instances active needed for the application to run
    - can be user defined
- Desired capacity
    - default is at min capacity if not specified
- Maximum capacity
    - sets maximum boundaries for scaling out

Directing Traffic with Elastic Load Balancing

Load Balancer ("Host")
- load balancer receives all incoming traffic and distributes the workload across multiple instances
- NOT an internal AWS tool, there are many pre-existing load balancers
    -

Elastic Load Balancing - "set and forget"
- High performance
- Cost-efficient
- Highly available
    - ELBs are regional constructs
    - runs on a region level, instead of on individual instances, so it's automatically highly available
- Automatically scalable
    - designed to handle additional traffic throughput with no additional cost

- ○ scales out - auto scaling service lets ELB know it's ready to handle traffic
- ○ scales in - stops new traffic and waits for existing requests to be complete
- ○ then extra instances are terminated
- ● Architecture
  - ○ Ordering Tier (front end) <-> Production Tier (back end)
  - ○ each tier shifts constantly based on command
  - ○ ELB handles interaction between ordering and production tier (very useful since there can be over thousands backend and frontend instances running across hundreds of tiers)

Messaging and Queuing

Monolithic applications and microservices
- ● Monolithic Application - Tightly Coupled Architecture
  - ○ when applications communicate directly
  - ○ if a single component fails or changes, it affects the entire system
- ● Microservice - Loosely Coupled Architecture
  - ○ single component failure won't cause the system-wide issue
  - ○ generally more reliable compared to tightly coupled
  - ○ introduction of a **message queue** (buffer) between applications
    - ■ if a component fails, the buffer will keep it from interacting with other functioning components, therefore

Amazon Simple Notification Service (Amazon SNS)
- ● A channel for messages to be delivered
- ● Publish Subscribe model (pub-sub model)
  - ○ configure subscribers to the topic, then find published messages for those subscribers (one topic, many subscribers)
  - ○ subscribers can be endpoints ("fan out" notifications)

Amazon Simple Queue Service (Amazon SQS)
- ● Allows us to send, store, and receive messages between software components at any volume
- ● Messages are stored in SQS queues until they are processed
  - ○ AWS manages the underlying architecture on how those queues are hosted
- ● Payload: data contained within the message

Additional Compute Services

Serverless computing
- ● we cannot see or access underlying architecture (it's taken care of for us!!)

AWS Lambda
- ● a serverless compute option
- ● code goes into "Lambda function" -> configure a trigger -> system waits for the trigger

- trigger is detected -> code is automatically run in a managed environment (completely taken care of by AWS)
- Runtime < 15 minutes (not for longer processes like deep learning)

Containers
- good for efficiency and portability
- Container orchestration tools
  - docker containers
  - container: package for the code, dependencies, any configurations
    - run on EC2 instances and in isolation from each other
    - host is an EC2 instance
  - cluster
    - multiple EC2 containers
    - it's hard to run multiple containers, so AWS helps with that
  - Amazon Elastic Container Service (Amazon ECS)
    - can run containerized applications at scale
  - Amazon Elastic Kubernetes Service (Amazon EKS)
    - does the same thing but with different tools
- AWs Fargate
  - serverless compute platform for ECS and EKS
  - basically just manages Amazon EC2 for you

	EC2	Serverless Computing (AWS Lambda)	Container orchestration tools (ECS / EKS)	AWS Fargate
full access to OS	x			
do not want to manage server		x		
want to run Docker container			x	
containers managed by me			x	
containers not managed by me			x	x

Module 3: Global Infrastructure and Reliability

AWS Global Infrastructure

Regions

- Regions: **geographically isolated** areas where you can access services needed to run your enterprise
- Running an application from a single location is generally a bad idea
- Regions are located closest to where business demands are highest
- Each region contains multiple data centers and is isolated from other regions
  - in order for data to be moved across regions, it must be specified by the user

Selecting a Region
- Compliance with data governance and legal requirements
  - companies may
- Proximity to customers
  - closer proximity to customers will help get content to them faster
  - a company based in one region can have a large customer base in another region, so they can run infrastructure at their own base, and run the applications close to their customers' region
- Available services within a region
  - Sometimes, the closest region might not have all the features you need
  - sometimes requires AWS to build out physical hardware one Region at a time
  - might need to run in the next closest region that already offers the feature you need
- Pricing
  - costs can vary from region to region, so the Region can be selected based on budget

Availability Zone
- Availability Zone (AZ): a single data center or group of data centers within a Region
  - Usually built far apart enough to avoid multiple AZ from being affected if one part of the Region was hit by a natural disaster, but close enough to have low latency
  - Recommended number of AZ: 2
- Elastic Load Balancer (ELB) mentioned!! this is what is means when they say it's a regional construct
- Regionally scoped service
- Benefits: High availability and disaster recovery

Edge Locations
- definition: a site that Amazon CloudFront uses to store cached copies of your content closer to your customers for faster delivery
- ContentDelivery Network (CDN): caching data to bring them closer to customers around the world
  - **Amazon CloudFront** is Amazon's CDN
- AWS Outposts
  - a mini-region inside a data center using 100% of AWS functionality but isolated inside your own building
  - good for when you want something contained on-premises, but you can still extend AWs infrastructure and services in other locations

How to Provision AWS Resources

Interacting with AWS Services: API calls
- AWS Management Console
  - browser based, visual management, good for starting out
  - good for test environments, view AWS bills, view monitoring, work with non-technical resources
- AWS Command Line Interface
  - can control multiple AWS services directly from the command line
  - automate actions through scripts
- Software Development Kits (SDKs)
  - allows AWS service API calls to be made using your programming language or platform
  - makes AWS services compatible with existing applications and can also allow you to create new applications

AWs Elastic Beanstalk
- service that helps provision Amazon EC2 environments
- deploys resources necessary to
  - adjust capacity
  - load balancing
  - automatic scaling
  - application health monitoring

AWS CloudFormation
- treats infrastructure as code - can build an environment by writing lines of code instead of using a management console
- provisions your resources in a safe, repeatable manner - minimizes manual action

Module 4: Networking

Connectivity to AWS

Amazon Virtual Private Cloud (Amazon VPC)
- A networking service that you can use to establish boundaries around AWS resources
- allows you to provision an isolated section of the AWS Cloud
- connection travels through the public internet, to a dedicated connect (unlike AWS direct connect)
- Within a VPC, you can organize resources into subnets
  - Subnet: a section of a VPC that can contain resources such as Amazon EC2 instances

Internet gateway
- Internet gateway: a connection between a VPC and the internet
- allows public traffic from the internet to access you VPC

Virtual private gateway

- Virtual private gateway: component that allows protected internet traffic to enter into the VPC
- acts like a bodyguard that encrypts internet traffic
- VPG allows you to establish a virtual private network (VPN) connection between your VPC and private network
- only allows in traffic from approved networks

AWS Direct Connect

- AWS Direct Connect: a service that lets you establish a dedicated private connection between your data center and a VPC
  - a corporate data center directs network traffic to an AWS Direct Connect location
  - traffic is then routed to a VPC through a virtual private gateway
  - all network traffic flows through this private connection
- helps reduce network costs and increase the amount of bandwidth that can travel through your network

Subnets and Network Access Control Lists

Subnets

- Public: contain resources that need to be publicly accessible (ex. online shop site)
- Private: contain resources that should only be accessible through your private network (ex. customer personal information)
- public and private subnets can communicate with each other!!

Network traffic in a VPC

- packet: unit of data sent over the internet or a network
  - data in an application hosted by AWS Cloud is sent to a customer in a packet when requested
- The packet enters the VPC through an internet gateway
- before the packet can enter or exit a subnet, it checks for permission
  - the permission contain info like who sent the packet and how the packet is trying to communicate with the resources in the subnet

Network ACLs (Access Control List)

- network ACL: virtual firewall that controls inbound and outbound traffic at the subnet level
- Default network ACL
  - allows all in/outbound traffic
- Custom network ACL
  - all in/out traffic is denied unless specified

Stateless packet filtering

- Does not save packet history, even if the packet has already moved across the subnet border previously
- Every single response has to checked and then approved/denied
- Security group: VPC component that checks packet permissions for an Amazon EC2 instance

Security groups

- security group: virtual firewall that controls in/outbound traffic for an EC2 instance

- by default
  - stateful
  - denies all incoming
  - allows all outgoing
- multiple EC2 instances can all be associated with the same security group or assigned different ones

Stateful packet filtering
- Saves previous decision for incoming packets
- When a packet response for a request returns to the instance, the security group remembers the previous request and allows the response to proceed

Global Networking

Domain Name System (DNS)
- DNS resolution: the process of translating a domain name to an IP address
  - the "phone book" of the internet
- DNS resolution involves a customer DNS resolver communicating with a company DNS server

Amazon Route 53
- Amazon Route 54: a DNS web service that gives developers and businesses a reliable way to route end users to internet applications hosted in AWS
- can manage the DNS records for domain names

Module 5: Storage and Databases

Instance Stores and Amazon Elastic block Store (Amazon EBS)

Instance Stores
- instance store: disk storage that provides block-level storage for an Amazon EC2 instance
  - physically attached to the most computer for an EC2 instance
  - has the same lifespan as the instance
  - data in instance store dies when the instance is terminated

Amazon EBS
- Amazon EBS: a service that provides block-level storage volumes that you can use with Amazon EC2 instances
- data in EBS does NOT die when the instance is terminated
- Configuration: define volume size and type and provision it
  - after creating an EBS volume, it can attach to Amazon EC2 instances
- Backing up is important because EBS is meant for data that needs to be persistent

Amazon EBS snapshots
- EBS snapshot: incremental backup
- first backup copies all the data -> subsequent backups only copies data that have changed

- different from full backups

Amazon Simple Storage Service (Amazon S3)

Object storage
- Object components: data, metadata, key
  - data: can be image, video, text document
  - metadata: what the data is, how it's used, size, etc
  - key: unique identifier

Storage classes
- S3 Standard
-

	Data access frequency	availability zones	speed	cost	summary
Standard	high	3+		high	wide range of use cases
Standard-IA	low	3+		low (storage) / high (retrieval)	infrequent access, high availability
One Zone-IA		1		low	
Intelligent-Tiering				mid	respondent, flexible
Glacier Instant Retrieval	high		fast	low	fast
Glacier Flexible Retrieval	mid		mid	low	archiving
Glacier Deep Archive	very low	3+	slow	lowest	long-term retention, infrequent access
Outposts			fast		on-premises, durable, across multiple devices

Selecting storage classes
- Two factors to consider when choosing storage class:
  - how **often** you plan on retrieving data
  - how **available** you need your data to be

Amazon Elastic File System (Amazon EFS)

File Storage

- multiple clients can access data stored in shared folders
- idea for large number of services/resources need to access the same data at the same time
- Amazon EFS: scalable file system used in AWS Cloud services and on-premises resources
  - changes size based on files added and removed, without interrupting applications

Comparing Amazon EBS and Amazon EFS

EBS	EFS
- single availability zone - in order to attach an EC2 instance to EBS volume, both instance and volume must be in the same AZ	- multiple availability zones - duplicate storage allows data to be accessed simultaneously across all AZ - uses AWS Direct Connect

Amazon Relational Database Service (Amazon RDS)

Relational database
- Relational database: database where data is stored in a way that relates it to other pieces of data
  - kind of like a dataframe
- use SQL -> easily understandable, consistent, scalable

Amazon Relational Database Service (RDS)
- Amazon RDS: service that enables you to run relational databases in the AWS Cloud
- manages automated tasks

Amazon RDS database engines
- six database engines
  - Amazon Aurora
  - PostgreSQL
  - MySQL
  - MariaDB
  - Oracle Database
  - Microsoft SQL Server

Amazon Aurora
- an enterprise-class relational database
- compatible with MySQL and PostgreSQL
- very fast!! (5x MySQL, 4x PostgreSQL)
- high availability, reliable, 3 AZ

Amazon DynamoDB

Nonrelational databases
- tables!!
- they use structures instead of rows of columns (no SQL involved)
- key-value pairs (okkk i see u being dictionary coded)

Amazon DynamoDB

- key-value database service
- crazy fast (single-digit millisecond performance at any scale)
- features
    - Serverless
        - do not have to provision, patch or manage servers
        - do not have to install, maintain, or operate software
    - Automatic scaling
        - scales capacity while data shrinks and grows, while maintaining consistent performance
        - high performance !!

Amazon Redshift

- Amazon Redshift: data warehousing service that you can use for big data analytics
- data comes from many sources
- helps understanding patterns in data

AWS Database Migration Service (DMS)

AWS DMS
- AWS DMS: service that enables you to move relational databases, non relational database, and other types of data stores
- source database <-data-> target database
    - source and target can be different for same types
    - source database remains operational during migration!!

Other use cases
- Development and test database migrations
    - testing apps with production data without affection production users
- Database consolidation
    - combining several databases into a single database
- Continuous replication
    - sending ongoing copies of your data to other target sources instead of doing a one-time migration

Additional Database Services
- DocumentDB
    - supports MongoDB workloads
- Neptune
    - graph database service
    - used to building and run apps with highly connected datasets (ex: recommendation engines)
- Quantum Ledger Database (QLDB)
    - ledger (record-keeping, usually financial) database

- - ○ can view complete history of changes to application data
  - Managed Blockchain
    - ○ service used to create and manage blockchain networks with open-source frameworks
    - ○ blockchain: distributed ledger system that lets multiple parties run transactions and share data without a central authority
  - ElastiCache
    - ○ a service that adds caching layers on top of your database to help improve the read times of common requests
    - ○ supports two data stores types: Redis and Memcached
  - DynamoDB Accelerator (DAX)
    - ○ an in-memory cache for DynamoDB
    - ○ helps improve response times from milliseconds to microseconds

Module 6: Security

AWS Shared Responsibility Model

Customer Responsibility **in** the Cloud
- responsible for everything they create and put in the cloud
- managing security requirements for the content
  - ○ what content you choose to store
  - ○ which AWS services
  - ○ who has access to the content
- factors
  - ○ complexity of your systems
  - ○ your company's specific operational security needs

AWS Responsibility **of** the Cloud
- operating, managing, and controlling all layers of infrastructure including
  - ○ host operating system
  - ○ virtualization layer
  - ○ physical security of data centers
- protecting global infrastructure that runs all AWS Cloud Services
- managing Cloud security
  - ○ Physical security of data centers
  - ○ hardware and software infrastructure
  - ○ network infrastructure
  - ○ virtualization infrastructure

User Permission and Access

AWS Identity and Access Management (IAM)
- AWS IAM: service that allows you to manage access to AWS services and resources securely

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

AWS account root user
- when you first create an AWS account, you start out as the **root user**
- has complete access to all AWS services and resources in the account
- Best practice
    - do NOT use for everyday tasks
    - instead, use it to assign permissions to create other users

IAM users
- identity that you create in AWS, representing the person or application that interacts with AWS
    - consists of name and credentials
- default: has no permissions
    - permission must be granted

IAM policies
- IAM policy: document that allows or denies permission to AWS services and resources
- enables you to customize user's levels of access
- Best practice
    - give out the "least privilege" when granting permissions
    - helps prevent users having more permission than needed for their tasks

IAM groups
- IAM group: a collection of IAM users
- policies assigned to a group will be assigned to all the users in the group
- policies at the group level makes it easier to adjust permissions when users transfer to a different role

IAM roles
- IAM role: an identity that lets you gain temporary access to permissions
- like groups, but more flexible
- good for when employees rotate through roles

Multi-factor authentication (MFA)
- added layer of security
- verifications are sent to a different device

AWS Organizations

AWS Organizations
- used to consolidate and manage multiple AWS accounts within a central location
- when you create an organization, AWS Organizations automatically creates a root
    - root - parent container for all the accounts in your organization
- Service Control Policies (SCPs): enable you to place restrictions on AWS services, resources, and individual API actions that users and roles in each account can access

- ○ can be applied to IAM users and an organizational unit

Organizational units (OU)
- ● OUs: groups of accounts to make it easier to manage accounts with similar business or security requirements

Compliance

AWS Artifact
- ● AWS Artifact: service that provides on-demand access to AWS security and compliance reports and select online agreements
- ● Two main sections:
  - ○ Agreements
    - ■ allows you to review, accept, and manage agreements for an individual account
  - ○ Reports
    - ■ provide compliance reports from third-party auditors
    - ■ they have tested and verified that AWS is compliant with a variety of global standards and regulations

Customer Compliance Center
- ● Customer Compliance Center: contains resources to help you learn more about AWS compliance
- ● can access compliance whitepapers and documentation about
  - ○ AWS answers to key compliance questions
  - ○ an overview of AWS risk and compliance
  - ○ an auditing security checklist
- ● includes an auditor learning path

Denial-of-Service Attacks

Denial-of-Service (DoS) Attacks
- ● DoS attack: deliberate attempt to make a website or application unavailable to users
- ● attacker can flood a website with excessive traffic until it's overloaded

Distributed Denial-of-Service Attacks
- ● DoS attacks coming from multiple sources, making it impossible to block them all and distinguish them from valid customer requests

AWS Shield
- ● AWS Shield: service that protects applications against DDoS attacks
- ● Two levels of protection
  - ○ Standard
    - ■ automatically protects all AWS customers at no cost
    - ■ protection against common types of DDoS attacks
  - ○ Advanced
    - ■ provides detailed attack diagnostics and ability to detect and mitigate complex DDoS attacks

- integrated with other AWS services like Amazon CloudFront, Amazon Route 53, EBL etc.
- can be integrated into AWS WAF by writing custom rules

Additional Security Services

AWS Key Management Service (AWS KMS)
- AWS KMS: enables you to perform **encryption operations** using cryptographic keys
- Cryptographic key: random string of digits used for locking and unlocking data
- AWS KMS can create, manage and use cryptographic keys
- can choose specific levels of access control that you need for you keys

AWS WAF
- AWS WAF: **web application firewall** that lets you monitor network requests that come into your web applications
- works with CloudFront and ALB (Application Load Balancer)
- AWS WAF blocks/allows traffic
    - using a web access control list (ACL)

Amazon Inspector
- performs **automated security assessments**
- Checks applications for security vulnerabilities and deviations from best practices
- after an assessment, Amazon Inspector makes a list of findings
    - list prioritizes by security level - description and a rec on how to fix it
- AWS does NOT guarantee that the rec works

Amazon GuardDuty
- Amazon GuardDuty: service that provides **intelligent threat detection** for AWS infrastructure and resources
- identifies threats by continuously monitoring the network activity and account behavior within your AWS environment
- after GuardDuty is enabled, it starts monitoring and you don't have to deploy or manage anything addition
- detailed reports on any detected threats can be viewed in the AWS Management Console

Module 7: Monitoring and Analytics
Amazon CloudWatch
Amazon CloudWatch
- Amazon CloudWatch: web service that enables you to monitor and manage different metrics and configure alarm actions based on metrics data
- uses metrics to represent data points for your resources
- AWS services send metrics to Cloudwatch
    - these metrics are then used to generate reports about performance over time
CloudWatch Alarms

- Alarms: automatically perform actions if the value of your metric has deviated significantly from a predefined threshold

CloudWatch dashboard
- CloudWatch dashboard: feature that allows you to access all metrics for your resources from a single location

AWS CloudTrail

- AWS CloudTrail: feature that records API calls for your account
    - info includes identity and source IP address of caller, time of call
- Can use API calls to provision, manage, and configure your AWS resources
- can view complete history of user activity and API calls

CloudTrail Insights
- CloudTrail Insights: optional feature that allows CloudTrail to automatically detect unusual API activities in your AWS account

AWS Trusted Advisor
- AWS Trusted Advisor: web service that inspects your AWS environment and provides real-time recs in accordance with AWS best practices
    - these inspections includes security checks (like Amazon S3 buckets with open access permissions)
- 5 best practice categories that Trusted Advisor compares its findings with:
    - cost optimization
    - performance
    - security
    - fault tolerance
    - service limits

AWS Trusted Advisor dashboard
- AWS Trusted Advisor dashboard: dashboard where the number of problems, investigations, and actions is indicated
    - can be access on the AWS Management Console
- For each category:
    - ✅ (green check): number of items with **no problems** detected
    - ⚠️ (orange triangle): number of **recommended investigations**
    - ⭕ (red circle): number of **recommended actions**

Module 8: Pricing and Support

AWS Free Tier
Three types of free offers:
- Always Free
    - do not expire and available to all AWS customers

- - ex: AWS Lambda allows 1 million free requests per month, Amazon DynamoDB allows 25GB of free storage per month
- 12 Months Free
  - free for a 12 month period after initial sign-up to AWS
  - ex: specific amounts of Amazon S3 Standard Storage
  - thresholds for monthly hours of Amazon EC2 compute time
- Trials
  - short term free trials start from the date you activate a particular service
  - different number of days depending on the service (Amazon Inspector offers 90 days, Amazon Lightsail offers 30 days)

AWS Pricing Concepts

How AWS Pricing works
- Pay for what you use
  - you are charged for exactly the amount of resources that you actually use
- Pay less when you reserve
  - Some services offer reservation options that provide significant discounts compared On-Demand Instance pricing
- Pay less with volume-based discounts when you use more
  - for services (ex. Amazon S3 storage price) with tiered pricing, per-unit cost goes down with increased usage AWS Pricing Calculator
- lets you explore AWS services and create an estimate for the cost of your use cases

Billing Dashboard

AWS Billing and Cost Management dashboard
- Compare current month balance with previous month, and get a forecast of next month
- view month-to-date spend by service
- View Free Tier usage by service
- Access Cost Explorer and create budgets
- Purchase and manage Savings Plans
- Publish AWS Cost and Usage Reports
  - AWS Cost and Usage Reports tracks AWS usage and provides estimated charges

Consolidated Billing

- Consolidated Billing: feature that enables you to receive a single bill for all AWS accounts for all AWS accounts in your organization
  - this feature is for **AWS Organizations**
- default max number of accounts per organization: 4 (but can be changed by contacting aWS Support)
- Benefits include

- - can easily track combined costs
    - can share bulk discount pricing, Saving Plans, and Reserved Instances across the organization

AWS Budgets

- AWS Budgets: service that allows you to create budgets to plan service usage, costs, and instance reservations
- updates 3 times a day
- you can set custom alerts to tell you when your usage exceeds the budget

AWS Cost Explorer

- AWS Cost Explorer: tool that lets you visualize, understand, and manage costs and usage over time
- includes default report for top 5 costing AWS services

AWS Support Plans

Basic Support
- free for all AWS customers !!
- includes access to whitepapers, documentation, and support communities
- can contact AWS for billing questions and service limit increases
- limited selection of AWS Trusted Advisor checks
- allows use of AWS PERsonal Health Dashboard

Developer
- access features including
    - Best practice guidance
    - client-side diagnostic tools
    - building-block architecture support (how to use AWS tools together)

Business
- access to additional features including
    - use-case guidance to identity optimal AWS tools
    - All AWS Trusted Advisor checks
    - Limited support for third-party software

Enterprise On-Ramp
- Has all the features of the above support plans
- additional accesses:
    - pool of Technical Account Managers
    - Cost Optimization workshop (one per year)
    - Concierge support team
    - Tools to monitor costs and performance through Trust Advisor and HEalth API/Dashboard
- Technical Account Manager pool services:

- - Consultative review and architecture guidance (one per year)
    - Infrastructure Event Management support (one per year)
    - Support automation workflows
    - 30 min or less response time for business-critical issues
  - Enterprise
    - has all the features of the ALL the above services
    - additional accesses:
      - **designated** Technical Account Manager
      - Operations Review and rolls to monitor health
      - Training and Game Days to drive innovation
- designated Technical Account Manager services:
  - Consultative review and architecture guidance (unlimited per year)
  - Infrastructure Event Management support (unlimited per year)
  - Cost Optimization Workshop and tools
  - 30 min or less response time for business-critical issues

Technical Account Manager (TAM)
- primary point of contact at AWS
- educates, empowers, and evolves your cloud journey across the full range of AWS services

AWS Marketplace

- AWS Marketplace: digital catalog that includes tons of software listings from indie software vendors
  - can be used to find, test, and buy software that runs on AWS
- Can explore software by industry and use case:
  - Infrastructure software
  - DevOps
  - Data PRoducts
  - Professional SErvices
  - Business Applications
  - Machine Learning
  - Industries
  - Internet of Things (IoT)

Module 9: Migration and Innovation

AWS Cloud Adoption Framework (AWS CAF)

Six Core Perspectives
- Business
  - **old (business and IT separate) -> new (business and IT integrated)**

- - ○ ensures that IT aligns with business needs and vice versa
    - ○ common roles: business and finance managers, budget owners, strategy stakeholders
- ● People
    - ○ evaluate organizational structures and roles, new skills, identify gaps
    - ○ prioritize **training, staffing**, and **organizational changes**
    - ○ common roles: human resources, staffing, people managers
- ● Governance
    - ○ focuses on the skills and process to align IT strategy with business strategy
    - ○ understand staff skills necessary to ensure **business governance in the cloud**
    - ○ common roles: Chief Information Officer (CIO), program managers, enterprise architects, business analysts, portfolio managers
- ● Platform
    - ○ principles and patterns for new cloud solutions
    - ○ common roles: Chief Technology Officer (CTO)
    - ○ IT managers
    - ○ Solutions architects
- ● Security
    - ○ ensures organization meets security objectives
    - ○ structure the selection and implementation of security controls that meet needs
    - ○ common roles: Chief Information Security Officer (CISO), IT security managers, IT security analysts
- ● Operations
    - ○ **run, use,** and **operate**, and recover IT workloads
    - ○ helps stakeholders define operations needed for successful cloud adoption
    - ○ common roles: IT operations managers, IT support managers

Migration Strategies

6 "R" Strategies
- ● Rehosting
    - ○ "lift-and-shift"
    - ○ moving apps without changes
    - ○ ex: company looking to migrate and scale quickly
- ● Replatforming
    - ○ "lift, tinker, and shift"
    - ○ optimizing the app by making a few cloud changes without changing the core architecture of the app
- ● Refactoring/re-architecting
    - ○ reimagining how an app is architectured by using cloud-native features
    - ○ ex: need to add features, scale, or performance that would be difficult to achieve in the app's existing env
- ● Repurchasing

- ○ tradition license -> software-as-a-service model
- ○ ex: business migrating from a customer relationship management (CRM) system to Salesforce.com (salesforce mentioned!!)
- ● Retaining
  - ○ keep applications that are critical for the business in the source environment
- ● Retiring
  - ○ removing apps that are no longer needed

AWS Snow Family
- ● physical devices to help physically transport data in/out of AWS

Family Members
- ● AWS Snowcone
  - ○ edge computing and data transfer device
  - ○ small, rugged, and secure
  - ○ 2 CPUs, 4GB memory, up to 14 TB storage
- ● AWS Snowball
  - ○ Snowball Edge **Storage** Optimized
    - ■ well suited for large-scale data migrations and recurring transfer workflows
    - ■ local computing
    - ■ higher capacity needs
    - ■ Storage (80 TB), Compute (40 vCPUs)
  - ○ Snowball Edge **Compute** Optimized
    - ■ well suited for resource heavy computing tasks
    - ■ ex: machine learning, full motion video analysis
    - ■ Storage (80 TB), Compute (104 vCPUs)
- ● AWS Snowmobile
  - ○ an exabyte-scale data transfer service used to move large amounts of data to AWS
  - ○ can transfer up to 100 petabytes of data per Snowmobile

Innovation with AWS
Innovate with AWS Services
- ● innovation in the cloud is driven by
  - ○ the current state
  - ○ the desired state
  - ○ problems needing to be solved
- ● Ways to Innovate
  - ○ Serverless applications
    - ■ serverless -> no need to provision, maintain, or administer servers (ex: AWS Lambda)
    - ■ allows you to focus on the core product !!
  - ○ Artificial intelligence
    - ■ AI powered services: Amazon Transcribe (speech to text), Amazon Comprehend (text pattern detection), Amazon Fraud Detector, Amazon Lex (chatbots)
  - ○ Machine learning

- - - ■ Amazon SageMaker: simplest ML process and lets you build, train, and deploy ML models quickly

Amazon Q Developer
- Amazon Q Developer: ML-powered code generator that provides code recs in real time
- Compatible with VSCode, PyCharm, Amazon SageMaker, Jupyter etc.
- Supports wide range of programming languages

Module 10: The Cloud Journey

The AWS Well-Architected Framework

- AWS Well-Architected Framework helps you understand how to design and operate reliable, secure, efficient, and cost-effective systems in AWS Cloud
- Six principles
  - Operational excellence
    - ability to run and monitor systems
    - ex: performing code, annotating documents, anticipating failure
  - Security
    - ability to protect information, systems, and assets while driving business value
    - automate best practices, security at all layers, protect data both in and out of transit
  - Reliability
    - ability to **recover** from disruption, dynamically **acquire computing resources** to meet demand, **mitigate** disruptions
    - ex: testing recovery procedures, scaling horizontally, automatically recovering from failure
  - Performance efficiency
    - ability to efficiently use computing resources to meet system requirements and adapt
    - ex: experimenting with architecture, going serverless, designing fast global systems
  - Cost optimization
    - ability to run systems to deliver business value at the lowest price point
    - ex: adopting a consumption model, analyzing expenditure, using managed services to reduce the cost of ownership
  - Sustainability
    - ability to reduce energy consumption and increasing efficiency across all components, minimizing total resources required
    - need to: understand impact, establish sustainability goals, use managed services, reduce downstream impact

Benefits of the AWS Cloud

- Trade upfront expense for variable expense
  - only pay for what you use
- Benefits from massive economies of scale
  - lower variable cost due to strong customer base
- Stop guessing capacity
  - no longer need to predict how much infrastructure capacity you need to deploy an application
  - only access resources you need
- Increase speed and ability
  - flexibility for development and deployment -> more time for experimentation and innovation
- Stop Spending money and maintaining data centers
  - less time spent on these tasks -> more time for apps and customers
- Go global in minutes
  - allows quick deployments globally with low latency

Module 11: Exam Preparation

Main concepts
- Cloud Concepts (weighted 24%)
- Security and Compliance (weighted 30%)
- Technology (weighted 34%)
- Billing and Pricing (weighted 12%)

Sample Exam questions

https://d1.awsstatic.com/training-and-certification/docs-cloud-practitioner/AWS-Certified-Cloud-Practitioner_Sample-Questions.pdf