Visualize the diagram of each problem

Practice Test:

- **Initial Practice Test: 59% (27/45)**
- **Overall Test 1: 80% (52/65)**
    - **77% (17/22)**
    - **79% (11/14)**
    - **80% (8/10)**
    - **84% (16/19)**
- **Overall Test 2: 74% (48/65)**
    - **73% (29/40)**
    - **90% (9/10)**
    - **67% (10/15)**
- **Overall Test 3: 75% (49/65)**
    - **65% (24/37)**
    - **89% (25/28)**
- **Overall Test 4:**
    - **75% (15/20)**
    - **94% (15/16)**

Mistakes

- Always make sure I am answering the question. There are usually many hints in the prompt.

AWS Notes

- Security group - virtual firewall to control inbound and outbound traffic. Operate on an instance level.
    - VPC has a default security group. If an instance isn't assigned a security group, it gets the default.
    - Use security groups for SSH connections to allow for data to return.
    - Updates to security groups take effect immediately
    - Backend security groups can have the frontend security group ID as the source to provide access to them.
- NACLs
    - Stateless. Can't determine between responses to inbound traffic and other outgoing outbound traffic.
    - The default network ACL allows all traffic to flow in and out of the subnets it is associated with.
- IAM
    - Identities: users, groups of users, roles
    - Policies: attached to identities or resource

        - when IAM policies are changed, they can impact the user experience of federated-login users
        - S3 bucket policies are an example of resource-based policies
    - Best practice: Roles can be used to assign access to users, applications, or services that don't normally have access to AWS resources
    - When you create an IAM user, you can specify console or programmatic access.
        - Programmatic includes both CLI and API.
- Cognito
    - Federated identity (Facebook, Google) - used to authenticate users
    - Then associate a Cognito identity pool to authorize users
- VPC
    - VPC Flow Logs - captures information about the IP traffic going to and from network interfaces in your VPC. Tracks flow of activity into and out of the VPC.
    - VPC Peering
        - VPC peering connects two VPCs in different regions
        - In VPC peering, using the NAT Gateway of another VPC is not supported
    - VPC Endpoints
        - S3, DynamoDB sit outside VPCs. Need to use VPC endpoints to access these resources.
        - Gateway endpoints
            - A gateway you specify in your route table for traffic destined to support AWS service
                - Currently supports S3 and DynamoDB
        - Interface endpoints
            - An Elastic Network Interface (ENI) with a private IP address that serves as an entry point for traffic destined to support a service. It enables you to privately access services by using private IP addresses.
    - Amazon Redshift Enhanced VPC Routing allows VPC resources access to Redshift
    - PrivateLink - VPC Endpoint Service
        - Provides secure private connectivity for services between separate VPCs.
        - Can use Network Load Balancer as well
    - Subnets
        - Each AZ will needs its own subnet
        - Public, Internet-facing instances should be in a different subnet than private or intranet applications
    - Bastion host provides inbound access to private instances from the public.
    - NAT Gateway provides outbound internet access from private instances
    - You can associate Secondary CIDR to your current CPV to accommodate more hosts
- S3
    - Strong read-after-write consistency

- Cross-region replication is a bucket-level configuration that enables automatic, asynchronous copying of objects across buckets in different AWS Regions.
- Lifecycle policies
  - Can be used to delete incomplete multipart file uploads.
  - Can delete content after a set period
- S3 event notifications
  - Can be sent to SNS topic, SQS queue, or a Lambda function
- Storage Types
  - Glacier - cheaper, archival storage.
    - Must use S3 Console to restore objects first prior to being used/copied.
    - Retrieval
      - Standard for low cost, 3-5 hours.
      - Expedited for faster retrieval.
      - Bulk for large amounts of data within 5-12 hours.
  - Standard
  - Infrequent Access
- With version-enabled S3 buckets, each version of an object can have a different retention period.
- Object ACL provides granular control on each file in the Amazon S3 bucket.
- Glacier Console
  - Use S3 Glacier Console to access objects/vaults of objects.
  - Use S3 Glacier Console Select to query specific continent data directly from Amazon S3 Glacier using simple SQL.
  - Use S3 Console to actually restore the objects.
- S3 Select
  - Can be used to query a subset of data from the objects stored in S3 buckets using simple SQL.
    - Objects must be stored using JSON, CSV, or Apache Parquet format.
      - GZIP and BZIP2 compression is supported for CSV, JSON.
- Uploads
  - Can use multi-part upload
- Downloads
  - Can use “range” HTTP header in a GET request to download specific bytes of an object
- Cross-origin Resource Sharing (CORS)
  - Allows websites to make requests to other websites
  - “AllowedMethod” supports GET, PUT, POST, DELETE, and HEAD
- Same-origin Policy
  - Scripts and other active content loaded from one site or domain cannot interfere or interact with content from another location with specific permission

- AWS KMS - Key Management Service used to create and control encryption keys

    - KMS master keys are region specific
    - Performs rotating automatically
- EC2
    - Types
        - On-demand
        - Spot instances (use spot pricing, at a discount to on-demand)
        - A cost effective method is to use a mix of on-demand and spot instance types
    - Can have either EBS or Instance Store volumes
    - Can create AMIs from EC2s. Can create EC2s from AMIs. Useful for launching EC2 instances.
        - After creating an AMI, create a launch configuration to launch the AMI.
    - Savings Plans - offers discount in exchange for a specific usage commitment for a one to three year period
        - Compute Savings Plan
            - Applies across EC2, Lambda, Fargate (can save up to 66%)
        - EC2 Instance Savings Plan
            - Only EC2 usage within the same instance family in the same region (can save up to 72%)
    - Placement groups
        - Clustered - low latency and high network throughput
        - Partition - useful for distributed and replicated workloads with some fault tolerance
        - Spread - most resilient
    - Hibernate
        - Launch an EC2 instance with an EBS root volume attached and enable hibernate.
        - When an EC2 is in hibernate, you only pay for the EBS volumes and elastic IP Addresses attached to it.
    - When you launch an EC2 instance, you have the option of passing “User Data” to the instance. You can perform common automated configuration tasks and even run scripts when the instance boots.
    - Public IPv4 addresses of EC2 Instances are released when an instance is restarted. They will need to be remapped.
        - Instances retain any private IPv4 addresses and any public IPv6 addresses.
- EBS
    - Storage types
        - Amazon EBS Cold HDD is for data accessed infrequently and the MOST cost-effective
        - Provisioned IOPS SSD - lowest latency, high throughput workloads
    - To make snapshots available in another region, create the snapshot and copy the snapshot to a new region.

- ECS
    - Tasks
        - Can use tasks to specify pretty much everything needed to run Docker containers
            - When you launch an ECS container instance, you have the option of passing “User Data” to the instance. You can perform common automated configuration tasks and even run scripts when the instance boots.
            - Most common use cases for user data are to pass configuration information to the Docker daemon, and the Amazon ECS container agent.
        - How to Docker
            - Create a docker image of you batch processing application
            - Deploy the image as an Amazon ECS task
- AWS Fargate
    - Run containerized applications without the need to provision and manage the backend infrastructure. Just register task definitions and Fargate launches the container for you.
- AWS Lambda
    - Stateless and very scalable.
    - Executes shorter bits of code.
    - Don’t need to do maintenance on the underlying infrastructure.
    - To write the logs from Lambda into AWS CloudWatch Logs
        1. CreateLogGroup
        2. CreateLogStream
        3. PutLogEvents (API) or put-log-events (CLI)
    - Don’t embed Access keys in a Lambda!
- API Gateway
    - An API management tool between a client and a collection of backend services.
    - Accepts API connections, aggregates the various services required to fulfill them, and returns the appropriate result.
- AWS Serverless Application Model (SAM) - open-source framework for building serverless applications in AWS.
- Auto Scaling
    - Auto Scaling Groups (ASG) - collections of EC2s
    - Scaling policies
        - Target tracking - targets an average aggregate CPU utilization of the ASG
        - Scheduled scaling - for situations where you know when and how much scaling you will need
    - Lifecycle hooks provide a specific amount of time (default is 1 hour) to complete a lifecycle action. After this time, the instance transitions to the next task.
        - Use Lifecycle Hooks to put the instances into a wait state before termination. Default wait period is 1 hour. During this time, you can perform custom activities to retrieve critical operational data.

- Elastic Beanstalk
    - For quickly deploying and scaling web applications. Can capacity provision, load balance, scale, and health monitor applications.
    - Can either handle HTTP requests or pulls tasks from an SQS queue
- DynamoDB (non relational)
    - Fully managed, serverless NoSQL database
    - Fast, predictable performance with seamless scalability
    - Stored in JSON format. Good for dynamic data.
    - Most efficient storage mechanism for just storing metadata that needs to be indexed
    - DynamoDB Streams - immediately after a change is made to a table, the event is added to a stream
    - Global tables can be used to make DynamoDB multi-region, multi-master.
- RDS (relational database) - more generic RDS service
    - You can only enable encryption when you create a database instance.
    - To encrypt a database
        1. Create a snapshot of an unencrypted database.
        2. Copy the snapshot and encrypt the new snapshot.
        3. Restore a new DB instance from the encrypted snapshot.
        4. (Optionally) Configure a Read Replica.
    - Automated backup process - snapshots are taken on a daily basis. Transaction logs are captured every 5 minutes.
    - Can create Read Replicas in different regions
        - Read replicas experience replication lag.
    - Multi-AZ databases aren't really needed in dev. Kinda expensive. Just save for production.
- Aurora (relational database) - enterprise-class MySQL and PostgreSQL database
    - 64TB limit. Up to 15 Replicas.
    - Works well for read scaling.
- Redshift - **Keywords: Data warehouse, Data mining**
    - Petabyte-scale data warehouse. Options for columnar storage.
    - Encryption - Use AWS KMS or HSM (Hardware security module) to manage encryption
    - Can use Reserved Instances for the Redshift Cluster
    - Automated snapshots are used for backups. Automatically deleted after 1-35 days.
- Amazon EMR - managed cluster platform for running big data frameworks, such as Apache Hadoop and Apache Spark.
- SNS
    - Service for sending emails.
- SQS
    - Message queueing service. Max payload size is 2GB.
    - Very horizontally scalable. Used to decouple systems.
    - For scaling write capacity for RDS, you can use SQS to queue messages.

    - For low and high priority customers, you can use two SQS queues and pull from the high priority queue first.
    - Dead-letter queues are for messages that can't be processed successfully. Useful for debugging your application.
- Kinesis
    - Massively scalable and durable real-time data streams that collect, process, and analyze data
    - Default retention period is 24 hours
    - Kinesis Data Firehose - loads streaming data in data lakes, data stores, and analytics services.
        - Can send to S3, Redshift, OpenSearch Service
        - Also generic HTTP endpoints, Datadog, New Relic, MongoDB, Splunk
- Amazon OpenSearch Service (successor to Amazon Elasticsearch Service)
    - Kinesis Data Firehose and CloudWatch Logs have built-in support for OpenSearch service
    - S3, Kinesis Data Streams, and DynamoDB can use Lambdas as event handlers
- CFTs - CloudFormation templates
        - Resources - defines the main AWS resources in the template
        - Parameters - parameters taken during template deployment
        - Outputs - return outputs after a CFT is executed
        - Mappings - map key-value pairs in a template
    - Drift Detection - can be used to detect changes made to AWS resources outside of the CloudFormation templates
        - Only monitors values explicitly set. (Doesn't monitor changes to default values). Explicitly set the property value if you want it monitored.
- CloudWatch - Application and Infrastructure Monitoring
    - Collects monitoring and operational data in the form of logs, metrics, and events
    - Can use CloudWatch to detect anomalies, set alarms, visualize logs and metrics, troubleshoot
        - Can also automate responses
    - CloudWatch Metrics
    - CloudWatch Logs
        - Log Stream - sequence of log events the share the same source
        - Log Group - group of log streams
        - PutLogEvents uploads batches of log events to CloudWatch Logs
    - To use CloudWatch Logs, you must allow the API Gateway access to CloudWatch Logs.
    - Execution logging - automatic by API Gateway
        - API Gateway manages the CloudWatch Logs
        - Creates log streams and log groups
        - All user requests and responses as well as error traces are logged
    - Access logging - done by user
        - Captures which resources accessed an API and the method used to access the API

- CloudTrail - Records and audits AWS account
    - Records all AWS account activity. Captures actions by users, roles, and AWS services
    - Enables governance, compliance, operating auditing, and risk auditing of your AWS account.
    - Logs stored in an S3 bucket are automatically encrypted with SSE.
        - Use CloudTrail log file integrity validation feature to make sure the logs are not tampered with.
    - Expensive!
- AWS Trusted Advisor
    - Provides recommendations for following AWS best practices
- CloudFront - Content Delivery Network (CDN)
    - Can cache frequently used static content in edge locations.
    - Can perform fast, basic “edge functions”
    - Origin Access Identity (OAI) - a special user that can control access to content in the Amazon S3 bucket.
        1. Create a special CloudFront user called an origin access identity.
        2. Associate it with your distribution.
        3. Configure S3 bucket permissions so that CloudFront can use the OAI to access the files in the bucket and serve them to users. Make sure users can’t use a direct URL to the S3 bucket to access a file there.
    - S3 Transfer Acceleration
    - Geo-restriction
        - Whilelist or blacklist
    - For third-party geolocation services, you can use signed URLs
    - Origin Protocol Policy - always set to “Match Viewer”
    - Viewer Protocol Policy - “Redirect HTTP to HTTPS” or “HTTPS only”
    - Use signed URLs or signed cookies to limit access.
        - Use signed URLs to restrict access to individual files or if the browser doesn’t support cookies
        - Use signed cookies to provide access to multiple files or you don’t want to change your URLs
    - CloudFront Events
        - Can log global service events to CloudTrail Trail.
            - Can have these events delivered to all regions. Can also have these events just delivered to a single region.
    - Query strings
        - Should always use the “&” character to delimit query strings.
        - Parameter names and values should use the same case.
        - Make sure the distribution is a web distribution.
- Route53 - Domain Name Service (DNS)
    - Translates URLs into IP addresses. IPv6 compliant.
    - Controls all the routing between resources.
    - Can register a domain with Route53.

    - Can host static websites in S3 and serve these.
    - Records
        - Amazon Route 53 records are standard DNS records
            - Points to IP address or domain name
        - Alias records provide a Route 53-specific extension to the DNS functionality
            - Can point to a CloudFront distribution, Elastic Beanstalk environment, a Load Balancer, an S3 bucket, or another Route 53 record
        - CNAME is similar to an Alias record but can only point to a subdomain whereas Alias records can point to both root domain and subdomains.
    - Routing
        - When you create a record, you choose a routing policy, which determines how AWS responds to queries.
        - Routing policies
            - Simple - use for a single resource that performs a function for your domain
            - Failover - when you want to configure active-passive failover
                - Can also failover to a static S3 bucket or CloudFront distribution
            - Geolocation (Geo DNS) - when you want to route traffic based on the location of your users
            - Geoproximity - when you want to route traffic based on the proximity of your resources
            - Latency - when you have resources on multiple AWS Regions and want to route traffic based on the lowest round-trip latency
            - Multivalue Answer - when you want to route traffic to up to eight healthy records at random
            - Weighted - route traffic proportionally
        - CNAME - an alias for domain names
        - Alias records automatically routes traffic to the new ELB IP addresses
- Regions
- AWS Global Accelerator
    - Provides two static IPs as endpoints
    - Directs traffic over AWS global network to nearest endpoint.
    - Improves availability and performance for applications.
- AWS Organizations
    - Create new accounts, allocate resources, group accounts, apply policies to accounts or groups, and simplify billing
    - To move the master account from one organization to another:
        - Remove all member accounts from the old organization
        - Delete the old organization
        - Invite the master account of the old organization to be a member or master account of the new organization

        - Consolidated Billing
            - A service for managing separate bills for every account and pay with the same account
        - Organizational Units (OU)
            - If a parent OU denies a policy, then it applies to all child OUs.
- AWS Resource Access Manager (RAM) to manage EC2 Dedicated Hosts centrally and share them with other member accounts
    - If resource sharing is not enabled for an organization, accounts can create a “resource share” and send an invitation to other accounts
    - Share the resource with AWS accounts of another organization in RAM
- AWS Directory Service
    - Enables end-users to use existing corporate credentials while accessing AWS applications
- AWS Managed Microsoft AD
    - Using VPN or Direct Connect, AWS Managed Microsoft AD can be used for both cloud and on-premise services

Gateways

- AWS Direct Connect Gateway
    - Can be used to establish high-performance network connections to different AWS Regions and reduce management loads.
    - Not a quick solution.
    - Does not provide IPSEC connectivity
- AWS VPN
    - Site-to-site VPN is the fastest and cost-effective way of establishing IPSEC connectivity from on-premise to AWS
- AWS Storage Gateway - Hybrid solution that offers unlimited AWS storage to an on-premise environment
    - File gateways - use file transfer protocols
        - S3 File
        - FSx File - high performance Windows File Server file shares
    - Volume-based gateways
        - Deployed into your on-premise environment as a VM running virtualization software
        - Two types - mount an iSCSI device to on-premise application servers
            - Cached - Frequent data is cached locally and everything else is stored in S3
            - Stored - all data is stored locally and backups are taken to S3
- Internet Gateway
    - Make sure the Route Table has an entry to the Internet Gateway
        - Add Destination 0.0.0.0/0 and Target as Internet Gateway

Migrations

- AWS Migration Hub - a service used to track and monitor the progress of migrations.
- AWS Data Sync
    - A service used to transfer data between on-premise storage to AWS S3, EFS, and FSx. Can transfer from NFS servers.
    - **Online** data transfers.
    - Can be bi-direcitonal.
    - Data verification (enable/disable)
- AWS Server Migration Service (SMS) - used to migrate servers + workloads to EC2
- AWS Database Migration Service - used to migrate databases from on-premise to AWS
    - Can keep the on-premise database fully operational during the migration.
    - Homogeneous migrations - use engine conversion tool
    - Heterogeneous migrations - use Schema Conversion Tool
- VM Import/Export
    - Allows customers to import VM images to create Amazon EC2 instances
- AWS Snowball - used to migrate terabytes to petabytes of data to S3
    - **Offline** data transfers

Random Services

- Athena is most suitable to run ad-hoc queries to analyze data in S3
    - Serverless, charged for the amount of scanned data
- Amazon Quicksight - visualizes data via dashboards
- Amazon SageMaker
    - Fully managed service for building, training, and deploying machine learning models.
- Amazon FSx - File Servers
    - FSx for Windows File Server
        - Can connect to Manage Active Directory and to Windows instances.
    - FSx for Lustre
        - An open-source parallel file system that can be used for high-performance computing (Machine Learning!).
        - Can read and write data into S3 and connect to multiple instances at the same time.
- AWS OpsWorks
    - Stacks
        - Collection of instances that are managed together for serving a common task.

Random

- Remote Desktop Protocol (RDP) - TCP/3389
    - SSH and RDP are the most common remote login options
- Bastion hosts sit in the public subnet (they need Internet access). They are used for external people to access things on the private subnet.

- NAT instances also sit in the public subnet. They are used by instances in the private subnet to initiate connections to access things on the open Internet.
    - Should launch NAT instances in multiple AZs and make them part of the Auto Scaling Group
- Use “query string parameters” to determine an action based on a URL
- Session data can be stored in ElastiCache or DynamoDB

Connection Types

- VPN - between on-premise and AWS resources
- VPC peering - between VPCs
- NAT Gateway - from private resources to the open Internet
    - Place at least one in each AZ
- Bastion host - from open Internet to private resources

Load Balancers

- Common features
    - Distribute requests to EC2 instances or Docker containers
    - Implement health checks and avoid unhealthy instances
    - Scale up and down
- ELB - also known as classic load balancer. Oldest type. Only balances traffic within a region.
- ALB - operates on Layer 7 (HTTP)
    - Can return a fixed response or redirection
    - Supports Server Name Indication (SNI), meaning it can server many domain names
        - Can attach 25 certificates
- NLB - operates on Layer 4 (TCP, UDP) - IP layer
    - High performance potential
    - Can be assigned elastic IP addresses
    - Real-time data streaming service (video, stock quotes) or non-HTTP protocols
- Other notes
    - API Gateway is typically better with a REST API architecture
    - Both Application Load Balancers and Network Load Balancers support dynamic mapping

CloudTrail

- **Monitors API activity across all AWS resources.** Can deliver log files from all regions to an Amazon S3 bucket. Records identity of the user, the start time of the AWS API call, the source IP address, the request parameters, and the response elements returned by the service

CloudFront is used for content distribution
ElastiCache can help with centralized bottlenecks

For AWS RDS:

- Multi-AZ allows high availability across AZs (but not regions)
- Read Replicas can be used to offload database reads, but not necessarily availability

Public subnets for web applications, Bastion Hosts, NAT Gateways
Private subnets for database layers

Pre-signed URLs for temporary access to S3 buckets.

- Can be better for restricting access to individual files.

Cookies are text files with small pieces of data - like a username or password - that are used to identify your computer.

- Cookies are better for providing access to multiple files (premium subscription)

Networking

- Elastic Fabric Adapter (EFA) - network device that you can attach to an EC2 instance to accelerate High Performance Computing (HPC) and machine learning applications
- Enhanced networking
    - Elastic Network Adapter (ENA)
    - Intel Virtual Function

Port 443 for HTTPS
Port 80 for HTTP

Resilient Architectures

- To move EBS volumes across regions, create EBS snapshots and then copy them to the destination region
- To move EC2 instances across regions, create AMIs for the underlying instances and copy them to the destination region
- Microservices architecture
    - Lambda - serverless compute
    - ECS - manage containers
    - API Gateway - serverless component for managing access to APIs
    - SQS - used for decoupling applications

Well-Architected Framework

- Monitoring and using alerts using CloudTrail and CloudWatch