AWS SAA – C02 Notes

Ajwad Javed
https://github.com/ajwadjaved

Dedicated Instance: your own instance on your own hardware
Dedicated Host: get access to the physical server itself, gives visibility to lower level hardware

To enable EC2 Hibernate, the EC2 Instance Root Volume type must be an EBS volume and must be encrypted to ensure the protection of sensitive content.
You have a critical application hosted on a fleet of EC2 instances in which you want to achieve maximum availability when there's an AZ failure. Which EC2 Placement Group should you choose? Spread Placement Group

- Cluster Placement Groups: A logical grouping of instances within a single AZ.
- Partition Placement Groups: Logical partition of instance groups such that no two partitions within a placement group share the same underlying hardware.
- Spread Placement Groups: each instance within a spread placement group will be placed in a different rack.

AMIs are startup templates (*AZ specific*!!) (note: AMI don't do shit about permissions). EBS Volumes is persistent storage, while snapshots are a backup of the EBS Volumes.
EBS volumes are network drives but with limited performance, EC2 instance store has better performance. Use for temporary data (buffer/cache).

In this scenario, the company has an existing IAM role hence you don't need to create a new one. IAM roles are global services that are available to all regions hence, all you have to do is assign the existing IAM role to the instance in the new region.

gp1/gp2 (this and below can only be used for boot volumes)
io1/io2 attach multi-attach (use clustered file system like GFS, not XFS/EX4 etc). *Multi-attach has to be across the same AZ* but can be with multiple EC2 instances.
st1/sc1 low/lowest cost HDDs (more for high throughput)

io1, is for > **30,000**
gp2, is for > 16,000
st1, is for > **500**
sc1, is for >250

- General Purpose SSD volumes (gp2 and gp3) balance price and performance for a wide variety of transactional workloads. These volumes are ideal for use cases such as boot volumes, medium-size single instance databases, and development and test environments.
- Provisioned IOPS SSD volumes (io1 and io2) are designed to meet the needs of I/O-intensive workloads that are sensitive to storage performance and consistency. They provide a consistent IOPS rate that you specify when you create the volume. This enables you to predictably scale to tens of thousands of IOPS per instance. Additionally, io2 volumes provide the highest levels of volume durability.

- Throughput Optimized HDD volumes (st1) provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing.
- Cold HDD volumes (sc1) provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential, cold-data workloads. If you require infrequent access to your data and are looking to save costs, these volumes provides inexpensive block storage.

EFS is a network file system (NFS) that allows you to mount the same file system on EC2 instances that are in different AZs. EFS vs EBS.

You are running a high-performance database that requires an IOPS of 310,000 for its underlying storage. What do you recommend?
You can run a database on an EC2 instance that uses an Instance Store, but you'll have a problem that the data will be lost if the EC2 instance is stopped (it can be restarted without problems). One solution is that you can set up a replication mechanism on another EC2 instance with an Instance Store to have a standby copy. Another solution is to set up backup mechanisms for your data. It's all up to you how you want to set up your architecture to validate your requirements. In this use case, it's around IOPS, so we have to choose an EC2 Instance Store.

Load Balancers usually send traffic to /health to check if other instances are up and running.

ASG auto-scaling groups, CloudWatch can set triggers for scaleIn scaleOut commands.

Good metrics to scale on: CPUUtilization, RequestCountPerTarget (outstanding requests), Average Network In/Ops (if application is network bound).

Only **Network Load Balancer provides both static DNS name and static IP**. While, **Application Load Balancer provides a static DNS name** but it does NOT provide a static IP. The reason being that AWS wants your Elastic Load Balancer to be accessible using a static endpoint, even if the underlying infrastructure that AWS manages changes.

"X-Forwarded-Port" used to get the client's requested port. "X-Forwarded-Proto" used to get the client's requested protocol. When using an Application Load Balancer to distribute traffic to your EC2 instances, the IP address you'll receive requests from will be the ALB's private IP addresses. To get the client's IP address, ALB adds an additional header called "X-Forwarded-For" contains the client's IP address.

Application Load Balancers support HTTP, HTTPS and WebSocket

Server Name Indication (SNI) allows you to expose multiple HTTPS applications each with its own TSL certificate on the same listener. Read more here:

https://aws.amazon.com/blogs/aws/new-application-load-balancer-sni/ (ALB can interact with multiple certificates using SNI)

A web application hosted on a fleet of EC2 instances managed by an Auto Scaling Group. You are exposing this application through an Application Load Balancer. Both the EC2 instances and the ALB are deployed on a VPC with the following CIDR 192.168.0.0/18. How do you configure the EC2 instances' security group to ensure only the ALB can access them on port 80?
Add an Inbound Rule for port 80 and ALB's Security Group as the source.
This is the most secure way of ensuring only the ALB can access the EC2 instances.
Referencing by security groups in rules is an extremely powerful rule and many questions at the exam rely on it. Make sure you fully master the concepts behind it!

Know what EFS is for and trust me, you don't need to know a lot of in-depth knowledge about it; just that it is for Linux instances, it can be accessed by lots of different instances, it is a Network File System and also that it is attached to a region.

RAID (redundant array of independent disks) is a storage technology that combines multiple disk drive components into a single logical unit so it behaves as one drive when connected to any other hardware. RAID 1 offers redundancy through mirroring, i.e., data is written identically to two drives. RAID 0 offers no redundancy and instead uses striping, i.e., data is split across all the drives. This means RAID 0 offers no fault tolerance; if any of the constituent drives fails, the RAID unit fails.

RDS
Master has to be encrypted for read replicas to be encrypted.
Multi-AZ keep the same SQL connection string with the M database. Adding more reader databases require individual connections so that we can load balance.

You work as a Solutions Architect for a gaming company. One of the games mandates that players are ranked in real-time based on their score. Your boss asked you to design then implement an effective and highly available solution to create a gaming leaderboard. What should you use?

- Use RDS for MySQL
- Use an Amazon Aurora
- Use ElastiCache for Memcached
- **Use ElastiCache for Redis - Sorted Sets**

You can not create encrypted Read Replicas from unencrypted RDS DB Instance. You can have **15** Aurora Read Replicas across a single DB Cluster upto 5 secondary regions (16 read replicas per secondary region).

Read Replica is async while Multi-AZ is synchronous.

Route53
Zone File contains DNS records. Name Server resolves DNS queries.

CNAME only routes non-root domain names.
Alias can route root and non-root domain names.

Q. You have purchased a domain on GoDaddy and would like to use Route 53 as the DNS Service Provider. What should you do to make this work?
A. Create public hosted zone and update 3rd party registrar NS records.
Public Hosted Zones are meant to be used for people requesting your website through the Internet. Finally, NS records must be updated on the 3rd party Registrar.

What is a DNS NS record? NS stands for 'nameserver,' and the nameserver record indicates which DNS server is authoritative for that domain (i.e. which server contains the actual DNS records). Basically, NS records tell the Internet where to go to find out a domain's IP address.

EC2 User Data allows you to automate the installation steps, but the installation would still take one hour to complete.
Golden AMI is an image that contains all your software installed and configured so that future EC2 instances can boot up quickly from that AMI.

Pre-signed URL: send credentials in URL.

S3 Bucket
SRR same region replication
CRR cross region replication

Can use S3 Batch Replication (replications won’t be chained, S1->S2->S3 won’t result in S1=S3)

Durability: how many times you can expect to lose a object (11 9’s)
Availability: how readily available a service is (s3 isn't available 53 mins a year)

S3 Transfer Acceleration: use edge locations to speed up parallel upload. Maximize private internet transfer and minimize public internet transfers.

S3 Select/Glacier Select -> SQL client side filtering to send filtered data to client.

Serverless SQL use Athena

Glacier Vault Lock (WORM write once read many times)

S3 Object Lock (governance mode need special privileges to edit, compliance mode no one can edit/change until time period -fixed/indefinite- is up)

Unicast IP (unique IPs)
Anycast IP (lowest latency, multiple devices have same IPs)

Signed Cookies are useful when you want to access multiple files. Signed URLs are for individual files.

Q. You are looking to get recommendations for S3 Lifecycle Rules. How can you analyze the optimal number of days to move objects between different storage tiers?
A. S3 Analytics

You can use **CloudWatch Logs** to monitor, store, and access your log files from EC2 instances, CloudTrail, Route 53, and other sources.

S3 Byte Range Fetch.

Q. You suspect that some of your employees try to access files in an S3 bucket that they don't have access to. How can you verify this is indeed the case without them noticing?
A. S3 Access Logs log all the requests made to S3 buckets and Amazon Athena can then be used to run serverless analytics on top of the log files.

Snowball cannot import directly to Glacier. Use Amazon S3 in combination with S3 lifecycle policy.

FSx
Scratch -> temporary, no persistence, temporary storage.
Choose HDD for high throughput.

SQS Temporary Queue Client -> request-response system (decouple request and response system)

End name with .fifo for FIFO queue

SNS Pub/Sub, (1) publishes a message and (n) can read/subscribe it from the published place

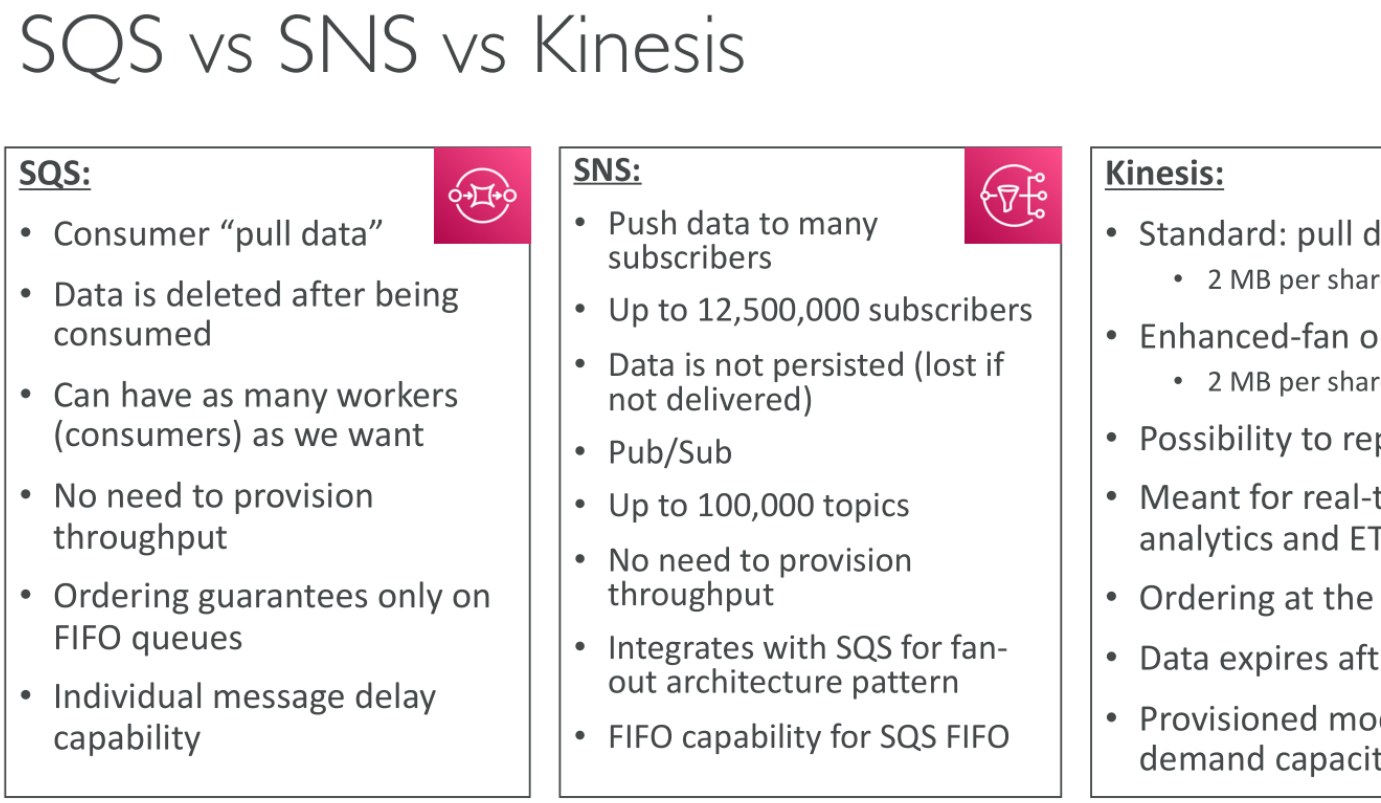
Kinesis Data Firehose destinations: redshift (copy through s3), s3, elasticsearch

SQS scales automatically. SNS doesn't.

This is a common pattern where only one message is sent to the SNS topic and then "fan-out" to multiple SQS queues. This approach has the following features: it's fully decoupled, no data loss, and you have the ability to add more SQS queues (more applications) over time.

Kinesis Data Stream uses the partition key associated with each data record to determine which shard a given data record belongs to. When you use the identity of each user as the partition key, this ensures the data for each user is ordered hence sent to the same shard.

SNS supported subscribers: Kinesis Data Firehose is now supported, but not Kinesis Data Streams. Also SQS, Lambda, HTTP endpoints are supported.



Long Polling
When a consumer requests messages from the queue, it can optionally "wait" for messages to arrive if there are none in the queue. LongPolling decreases the number of API calls made to SQS while increasing the efficiency and reducing latency of your application

With short polling, the ReceiveMessage request queries only a subset of the servers (based on a weighted random distribution) to find messages that are available to include in the response. Amazon SQS sends the response right away, even if the query found no messages.

With long polling, the ReceiveMessage request queries all of the servers for messages. Amazon SQS sends a response after it collects at least one available message, up to the maximum

number of messages specified in the request. Amazon SQS sends an empty response only if the polling wait time expires.

Amazon Elastic Container Registry (ECS) - store containers
ECR - store container images
EKR - kubernetes

Amazon ECR is a fully managed container registry that makes it easy to store, manage, share, and deploy your container images. It won't help in running your Docker-based applications.

Use EFS for persistent storage, can read between EC2 & ECS and no AZ issues.
(note: fsc for lustre and s3 storage not supported for ECS)

Methods to scale on:

- CPU
- RAM
- Request Count per Target (coming from ALB)

- Target Tracking, scale based on target value for specific cloudwatch metric
- Step Scaling, scale based on cloudwatch alarm
- Scheduled Scaling

Cloudwatch alarm can trigger ASG to setup more EC2 instances.

Amazon Eventbridge can trigger our task on ECS to do some serverless processing.

ECS Tasks = EKS Pods (latter is cloud agnostic)

Amazon Elastic Container Service (ECS) has two Launch Types: ..... and .....

EC2 Launch Type and Fargate Launch Type. (EKS is agnostic)

ECS Task Role is the IAM Role used by the ECS task itself. Use when your container wants to call other AWS services like S3, SQS, etc.

Q. You are deploying an application on an ECS Cluster made of EC2 instances. Currently, the cluster is hosting one application that is issuing API calls to DynamoDB successfully. Upon adding a second application, which issues API calls to S3, you are getting authorization issues. What should you do to resolve the problem and ensure proper security?
A. Create an IAM Task Role for the application

Q. Which feature allows an Application Load Balancer to redirect traffic to multiple ECS Tasks running on the same ECS Container instance?
A. Dynamic Port Mapping

Lambda Limits
Execution:
• Memory allocation: 128 MB – 10GB (1 MB increments)
• Maximum execution time: 900 seconds (15 minutes)
• Environment variables (4 KB)
• Disk capacity in the “function container” (in /tmp): 512 MB
• Concurrency executions: 1000 (can be increased)
Deployment:
• Lambda function deployment size (compressed .zip): 50 MB
• Size of uncompressed deployment (code + dependencies): 250 MB
• Can use the /tmp directory to load other files at startup
• Size of environment variables: 4 KB

DynamoDB, provisioned mode. Read Capacity Units, Write Capacity Units can be autoscaled but for on-demand mode no capacity planning is needed (2/3x more expensive).

AWS DAX -> dynamoDB cache

API Gateway – Security – Summary

- IAM:
  - Great for users / roles already within your AWS account
  - Handle authentication + authorization
  - Leverages Sig v4
- Custom Authorizer:
  - Great for 3rd party tokens
  - Very flexible in terms of what IAM policy is returned
  - Handle Authentication + Authorization
  - Pay per Lambda invocation
- Cognito User Pool:
  - You manage your own user pool (can be backed by Facebook, Google login etc…
  - No need to write any custom code
  - Must implement authorization in the backend

Custom Authorizer can be cached so you don;t have to invoke Lambda Authorizer each time.

An Edge-Optimized API Gateway is best for geographically distributed clients. API requests are routed to the nearest CloudFront Edge Location which improves latency. The API Gateway still lives in one AWS Region.

Amazon Cognito User Pools integrate with Facebook to provide authenticated logins for your application users.
Lambda@Edge is a feature of CloudFront that lets you run code closer to your users, which improves performance and reduces latency.

OLTP and OLAP: The two terms look similar but refer to different kinds of systems. Online transaction processing (OLTP) captures, stores, and processes data from transactions in real time. Online analytical processing (OLAP) uses complex queries to analyze aggregated historical data from OLTP systems.

Loading data into Redshift

- Kinesis Data Firehose automatically using S3 copy to send to Redshift Cluster
- Manually copying from S3 to Redshift Cluster
- EC2 instance batch write job to Redshift Cluster

Amazon Aurora is a MySQL and PostgreSQL-compatible relational database. It features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 128TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across 3 AZs.

Q Which feature in Redshift forces all COPY and UNLOAD traffic moving between your cluster and data repositories through your VPCs?
A Enhanced VPC Routing

The AWS Serverless Application Model (SAM) is an open-source framework for building serverless applications. It provides shorthand syntax to express functions, APIs, databases, and event source mappings. With just a few lines per resource, you can define the application you want and model it using YAML. During deployment, SAM transforms and expands the SAM syntax into AWS CloudFormation syntax, enabling you to build serverless applications faster. (Beanstalk for serverless)

Cloudwatch Unified Agent sends logs to Cloudwatch Logs

CloudWatch alarms -> EC2 modification, SNS message, ASG scaling

CloudTrail: management events, data events, insight events

CloudWatch vs CloudTrail vs Config

- CloudWatch
  - Performance monitoring (metrics, CPU, network, etc…) & dashboards
  - Events & Alerting
  - Log Aggregation & Analysis
- CloudTrail
  - Record API calls made within your Account by everyone
  - Can define trails for specific resources
  - Global Service
- Config
  - Record configuration changes
  - Evaluate resources against compliance rules
  - Get timeline of changes and compliance

If you set an alarm on a high-resolution metric, you can specify a high-resolution alarm with a period of 10 seconds or 30 seconds, or you can set a regular alarm with a period of any multiple of 60 seconds.
For EC2 high resolution is 1 second.

You would like to evaluate the compliance of your resource's configurations over time. Which AWS service will you choose?
AWS Config

You have enabled AWS Config to monitor Security Groups if there's unrestricted SSH access to any of your EC2 instances. Which AWS Config feature can you use to automatically re-configure your Security Groups to their correct state?
AWS Config Remediations

EC2 Detailed Monitoring -> when you want less than 5 minutes analytics.

Symmetric encryption -> private key
Asymmetric encryption -> public/private key

If data >4KB then envelope encryption (else KMS).

If want to copy snapshots then have to enable cross-account access for KMS.

Shield Advanced -> 24/7 support team
AWS Inspector is only for EC2 instances and container infrastructure.

AD Connector a proxy.
Simple AD standalone.
AWS Managed Microsoft AD supports MFA.

Service Control Policies **SCP**: apply to slave accounts in organizational units. Can completely block out some IAM actions (precedence works, rule applied by master take precedence).

To change master accounts you first have ot migrate each slave account, if a slave account has to change organizations it first has to leave it’s current organization and then accept an invite to the second one.

IAM Role vs Resource Based Policies (latter the principal doesn’t have to give up any of his permissions. Else you take the permissions assigned by the role otherwise).

Amazon Cognito can be used to federate mobile user accounts and provide them with their own IAM permissions, so they can be able to access their own personal space in the S3 bucket.

SAML Identity Federation is used to integrate an Identity Provider service such as Microsoft Active Directory with AWS. It does not work for mobile applications.

STS gives temporary access.

KMS Key Policies maintain permissions of the keys.

SSM Parameters Store can be used to store secrets and has built-in version tracking capability. Each time you edit the value of a parameter, SSM Parameter Store creates a new version of the parameter and retains the previous versions. You can view the details, including the values, of all versions in a parameter's history.

You would like to externally maintain the configuration values of your main database, to be picked up at runtime by your application. What's the best place to store them to maintain control and version history?
SSM Parameter Store.

AWS Guarduty doesn’t scan CloudWatch. Does do CloudTrail, VPC Flow Logs, DNS Logs.

WAF protects against web app attacks like SQL injections (HTTP level hence the firewall).

AWS Firewall Manager is a security management service that allows you to centrally configure and manage firewall rules across your accounts and applications in AWS Organizations. It is integrated with AWS Organizations so you can enable AWS WAF rules, AWS Shield Advanced protection, security groups, AWS Network Firewall rules, and Amazon Route 53 Resolver DNS Firewall rules.

Amazon Macie is a fully managed data security service that uses Machine Learning to discover and protect your sensitive data stored in S3 buckets. It automatically provides an inventory of S3 buckets including a list of unencrypted buckets, publicly accessible buckets, and buckets shared with other AWS accounts. It allows you to identify and alert you to sensitive data, such as Personally Identifiable Information (PII).

The option that says: Enable CloudTrail Insights events is incorrect. CloudTrail Insights events is just an optional feature that allows you to detect unusual write API activities in your account.

VPC
0.0.0.0/0 means all IPs

For Exam subnet, calculate with 2^(32-x) then minus 5 to see if number is still >

Reachability Analyzer doesn’t send packets, just checks configs to see up/down links.

NACL

These are stateless, meaning any change applied to an incoming rule isn't automatically applied to an outgoing rule. These are stateful, which means any changes which are applied to an incoming rule is automatically applied to a rule which is outgoing.

Security Group vs. NACLs

Security Group	NACL
Operates at the instance level	Operates at the subnet level
Supports allow rules only	Supports allow rules and deny rules
Stateful: return traffic is automatically allowed, regardless of any rules	**Stateless:** return traffic must be explicitly a rules (think of ephemeral ports)
All rules are evaluated before deciding whether to allow traffic	Rules are evaluated in order (lowest to hig deciding whether to allow traffic, first mat
Applies to an EC2 instance when specified by someone	Automatically applies to all EC2 instances i subnet that it's associated with

NACL Examples: https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.htm

ENI - private ID address, supports most aws services
Gateway endpoint supports s3/dynamoDB

nat gateway, let private net talk to public

Site to site VPN: enable route propagation for virtual private gateway in the route table.
Transit Gateway alone supports IP multicast
ECMP Transit Gateway increases bandwidth of site to site VPN (equal cost multi path routing)

Egress is ipv6 only NAT gateway

VPN Peer, route tables have to be updated

These two services (s3/dynamoDB) have a VPC Gateway Endpoint (remember it), all the other ones have an Interface endpoint (powered by Private Link - means a private IP).

Site to Site VPN connection, you need to configure customer gateway and virtual private gateway

AWS VPN CloudHub allows you to securely communicate with multiple sites using AWS VPN. It operates on a simple hub-and-spoke model that you can use with or without a VPC.

If only ipv6 left then make additional CIDR ipv4

RPO -> how much data you lost
RTO -> downtime

backup and restore
pilot light - have database running but not server, can failover with route53
multi-site - have ec2 and database running but asg it running at minimal
hot site

SCT - scheme conversion for when 2 underlying tech different

PITR - point in time recovery

AWS CodePipeline - codecommit codebuild elastic beanstalk/codedeploy

Stacksets, create update delete multiple stacks across multiple accounts/regions

Step Functions (SWS is for when you need external signals and when slave returns any value back to parent process)

eliminate management of on-premise VDI (virtual desktop infrastructure) -> aws workspaces

AppSync (graphQL) vs cognito sync

Cloud Formation is IaC

Amazon ECS is a fully managed container orchestration service that helps you easily deploy, manage, and scale containerized applications.

AWS CodeDeploy is a fully managed deployment service that automates software deployments to a variety of computing services such as EC2, Fargate, Lambda, and your on-premises servers. You can define the strategy you want to execute such as in-place or blue/green deployments.

AWS Simple Workflow Service SWF helps you build, run, and scale background jobs that have parallel or sequential steps. It makes it easy to build applications that coordinate work across

distributed components. It is an old service, to orchestrate Lambda functions use AWS Step Functions instead.

AWS Step Functions is a low-code visual workflow service used to orchestrate AWS services, automate business processes, and build Serverless applications. It manages failures, retries, parallelization, service integrations-

use amazon sts for temp credentials, never store credentails locally
dynamo -> DAX caching layer
—

Cluster – packs instances close together inside an Availability Zone. This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.

Partition – spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.

Spread – strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

Blue/green deployment is a technique for releasing applications by shifting traffic between two identical environments running different versions of the application: "Blue" is the currently running version and "green" the new version. This type of deployment allows you to test features in the green environment without impacting the currently running version of your application. When you’re satisfied that the green version is working properly, you can gradually reroute the traffic from the old blue environment to the new green environment. Blue/green deployments can mitigate common risks associated with deploying software, such as downtime and rollback capability.

AWS Global Accelerator relies on ELB to provide the traditional load balancing features such as support for internal and non-AWS endpoints, pre-warming, and Layer 7 routing. However, **while ELB provides load balancing within one Region, AWS Global Accelerator provides traffic management across multiple Regions.**

Amazon Macie is a fully managed data security and data privacy service that uses machine learning and pattern matching to discover and protect your sensitive data on Amazon S3. Macie automatically detects a large and growing list of sensitive data types, including personally identifiable information (PII) such as names, addresses, and credit card numbers. It also gives you constant visibility of the data security and data privacy of your data stored in Amazon S3.

Use Amazon GuardDuty to monitor any malicious activity on data stored in S3. Use Amazon Macie to identify any sensitive data stored on S3

VPC sharing can only share subnets with different accounts, not the account itself.

Use API Gateway Lambda authorizer - If you have an existing Identity Provider (IdP), you can use an API Gateway Lambda authorizer to invoke a Lambda function to authenticate/validate a given user against your IdP. You can use a Lambda authorizer for custom validation logic based on identity metadata.

A Lambda authorizer can send additional information derived from a bearer token or request context values to your backend service. For example, the authorizer can return a map containing user IDs, user names, and scope. By using Lambda authorizers, your backend does not need to map authorization tokens to user-centric data, allowing you to limit the exposure of such information to just the authorization function.

When using Lambda authorizers, AWS strictly advises against passing credentials or any sort of sensitive data via query string parameters or headers, so this is not as secure as using Cognito User Pools.

In addition, both these options do not offer built-in user management.

Dedicated instances may share hardware with other instances from the same AWS account that are not dedicated instances. Dedicated instances cannot be used for existing server-bound software licenses.
Neither on-demand instances nor reserved instances can be used for existing server-bound software licenses.

Have to get dedicated host for detailed options.

RDS has a enhanced monitoring mode.

Amazon Inspector is for checking network availability of different instances/services in your infrastructure.

It is true that AWS WAF can filter web requests based on IP addresses, HTTP headers, HTTP body, or URI strings, to block common attack patterns, such as SQL injection or cross-site scripting. NACL, on the other hand, acts like a firewall for controlling traffic in and out of your subnets.

If the scenario is more about protecting your application from common web exploits (SQL injection or cross-site scripting), then AWS WAF would be a more suitable choice. Otherwise,

you should choose NACL if it explicitly requires the need to block all traffic based on a given IP address or range.

If you're using messaging with existing applications and want to move your messaging service to the cloud quickly and easily, it is recommended that you consider **Amazon MQ**. It supports industry-standard APIs and protocols so you can switch from any standards-based message broker to Amazon MQ without rewriting the messaging code in your applications.
If you are building brand new applications in the cloud, then it is highly recommended that you consider Amazon SQS and Amazon SNS. Amazon SQS and SNS are lightweight, fully managed message queue and topic services that scale almost infinitely and provide simple, easy-to-use APIs. You can use Amazon SQS and SNS to decouple and scale microservices, distributed systems, and serverless applications, and improve reliability.

The option that says: Migrate the existing file share configuration to **AWS Storage Gateway** is incorrect because AWS Storage Gateway is primarily used to integrate your on-premises network to AWS but not for migrating your applications. Using a file share in Storage Gateway implies that you will still keep your on-premises systems, and not entirely migrate it.

Enabling cross-account access is incorrect because cross-account access is a feature in IAM and not in Amazon S3.
Enabling Cross-Zone Load Balancing is incorrect because Cross-Zone Load Balancing is only used in ELB and not in S3.
Enabling Cross-Region Replication (CRR) is incorrect because CRR is a bucket-level configuration that enables automatic, asynchronous copying of objects across buckets in different AWS Regions.

The option that says: Use a RAID 0 storage configuration that stripes multiple Amazon EBS volumes together to store the files. Configure the Amazon Data Lifecycle Manager (DLM) to schedule snapshots of the volumes after 2 years is incorrect because RAID (Redundant Array of Independent Disks) is just a data storage virtualization technology that combines multiple storage devices to achieve higher performance or data durability. RAID 0 can stripe multiple volumes together for greater I/O performance than you can achieve with a single volume. On the other hand, RAID 1 can mirror two volumes together to achieve on-instance redundancy.

AWS Resource Access Manager (RAM) is a service that enables you to easily and securely share AWS resources with any AWS account or within your AWS Organization. You can share AWS Transit Gateways, Subnets, AWS License Manager configurations, and Amazon Route 53 Resolver rules resources with RAM.

Amazon API Gateway provides throttling at multiple levels including global and by service call. Throttling limits can be set for standard rates and bursts. For example, API owners can set a rate limit of 1,000 requests per second for a specific method in their REST APIs, and also configure Amazon API Gateway to handle a burst of 2,000 requests per second for a few seconds. **Amazon API Gateway tracks the number of requests per second. Any request**

over the limit will receive a 429 HTTP response. The client SDKs generated by Amazon API Gateway retry calls automatically when met with this response. Hence, enabling throttling limits and result caching in API Gateway is the correct answer.
You can add **caching to API calls by provisioning an Amazon API Gateway cache** and specifying its size in gigabytes. The cache is provisioned for a specific stage of your APIs. This improves performance and reduces the traffic sent to your back end. Cache settings allow you to control the way the cache key is built and the time-to-live (TTL) of the data stored for each method. Amazon API Gateway also exposes management APIs that help you invalidate the cache for each stage.

The number that follows the backslash represents the amount of bits that are blocked when defining the range. **A CIDR block of /0 would allow access to any IP address between 0.0. 0.0 and 255.255. 255.255, while a CIDR block of /32 would only allow access to the IP address that precedes it.**

In the given scenario, you can use Lambda@Edge to allow your Lambda functions to customize the content that CloudFront delivers and to execute the authentication process in AWS locations closer to the users. In addition, you can set up an origin failover by creating an origin group with two origins with one as the primary origin and the other as the second origin which CloudFront automatically switches to when the primary origin fails. This will alleviate the occasional HTTP 504 errors that users are experiencing. Therefore, the correct answers are:

Deploying Multi-AZ in API Gateway with Read Replica is incorrect because RDS has Multi-AZ and Read Replica capabilities, and not API Gateway.

Deploying Multi-AZ in API Gateway with Read Replica is incorrect because RDS has Multi-AZ and Read Replica capabilities, and not API Gateway.

AWS Lake Formation is a service that makes it easy to set up a secure data lake in days. A data lake is a centralized, curated, and secured repository that stores all your data, both in its original form and prepared for analysis. A data lake enables you to break down data silos and combine different types of analytics to gain insights and guide better business decisions.
AWS Lake Formation is integrated with AWS Glue which you can use to create a data catalog that describes available datasets and their appropriate business applications. Lake Formation lets you define policies and control data access with simple “grant and revoke permissions to data” sets at granular levels. You can assign permissions to IAM users, roles, groups, and Active Directory users using federation. You specify permissions on catalog objects (like tables and columns) rather than on buckets and objects.

Take note that there are certain differences between CloudWatch and Enhanced Monitoring Metrics. CloudWatch gathers metrics about CPU utilization from the hypervisor for a DB instance, and Enhanced Monitoring gathers its metrics from an agent on the instance. As a result, you might find differences between the measurements, because the hypervisor layer

performs a small amount of work. Hence, enabling **Enhanced Monitoring in RDS is the correct answer** in this specific scenario.
The differences can be greater if your DB instances use smaller instance classes, because then there are likely more virtual machines (VMs) that are managed by the hypervisor layer on a single physical instance. Enhanced Monitoring metrics are useful when you want to see how different processes or threads on a DB instance use the CPU.

Take note that a non-Serverless DB cluster for Aurora is called a provisioned DB cluster.

Using Amazon FSx For Lustre and Amazon EBS Provisioned IOPS SSD (io1) volumes for hot and cold storage respectively is incorrect because the Provisioned IOPS SSD (io1) volumes are designed for storing hot data (data that are frequently accessed) used in I/O-intensive workloads. EBS has a storage option called “Cold HDD,” but due to its price, it is not ideal for data archiving. EBS Cold HDD is much more expensive than Amazon S3 Glacier / Glacier Deep Archive and is often utilized in applications where sequential cold data is read less frequently.
USE S3/Lustre for cold/hot storage

The gateway provides access to objects in S3 as files or file share mount points. With a file gateway, you can do the following
– You can store and retrieve files directly using the NFS version 3 or 4.1 protocol.
– You can store and retrieve files directly using the SMB file system version, 2 and 3 protocol.
– You can access your data directly in Amazon S3 from any AWS Cloud application or service.
– You can manage your Amazon S3 data using lifecycle policies, cross-region replication, and versioning. You can think of a file gateway as a file system mount on S3.

With **short polling**, the ReceiveMessage request queries only a subset of the servers (based on a weighted random distribution) to find messages that are available to include in the response. Amazon SQS sends the response right away, even if the query found no messages.
With **long polling**, the ReceiveMessage request queries all of the servers for messages. Amazon SQS sends a response after it collects at least one available message, up to the maximum number of messages specified in the request. Amazon SQS sends an empty response only if the polling wait time expires.

The option that says: The web application is set for long polling so the messages are being sent twice is incorrect because long polling helps reduce the cost of using SQS by eliminating the number of empty responses (when there are no messages available for a ReceiveMessage request) and false empty responses (when messages are available but aren’t included in a response). Messages being sent twice in an SQS queue configured with long polling is quite unlikely.
The option that says: The web application is set to short polling so some messages are not being picked up is incorrect since you are receiving emails from SNS where messages are certainly being processed. Following the scenario, messages not being picked up won’t result into 20 messages being sent to your inbox.

The option that says: The web application does not have permission to consume messages in the SQS queue is incorrect because not having the correct permissions would have resulted in a different response. The scenario says that messages were properly processed but there were over 20 messages that were sent, hence, there is no problem with the accessing the queue.

In Amazon SQS, you can configure **the message retention period to a value from 1 minute to 14 days**. The default is 4 days. Once the message retention limit is reached, your messages are automatically deleted.
A single Amazon SQS message queue can contain an unlimited number of messages. However, there is a 120,000 limit for the number of inflight messages for a standard queue and 20,000 for a FIFO queue. Messages are inflight after they have been received from the queue by a consuming component, but have not yet been deleted from the queue.

Elastic Fabric Adapter (EFA) is a network interface for Amazon EC2 instances that enables customers to run applications requiring high levels of inter-node communications at scale on AWS.

Latency Routing lets Amazon Route 53 serve user requests from the AWS Region that provides the lowest latency. It does not, however, guarantee that users in the same geographic region will be served from the same location.
Geoproximity Routing lets Amazon Route 53 route traffic to your resources based on the geographic location of your users and your resources. You can also optionally choose to route more traffic or less to a given resource by specifying a value, known as a bias. A bias expands or shrinks the size of the geographic region from which traffic is routed to a resource.
Geolocation Routing lets you choose the resources that serve your traffic based on the geographic location of your users, meaning the location that DNS queries originate from.
Weighted Routing lets you associate multiple resources with a single domain name (tutorialsdojo.com) or subdomain name (subdomain.tutorialsdojo.com) and choose how much traffic is routed to each resource.

Although you can copy data from on-premises to AWS with **Storage Gateway**, it is not suitable for transferring large sets of data to AWS. Storage Gateway is mainly used in providing **low-latency access** to data by caching frequently accessed data on-premises while storing archive data securely and durably in Amazon cloud storage services. Storage Gateway optimizes data transfer to AWS by sending only changed data and compressing data.

With AWS DataSync, **you can transfer data from on-premises directly to Amazon S3 Glacier Deep Archive**. You don't have to configure the S3 lifecycle policy and wait for 30 days to move the data to Glacier Deep Archive.

sc1 < st1 (sc1 is even slower/st1 supports frequently accessed data). C come before T so C slower. C for cold storage while T for hot storage.

Elastic Load Balancers distribute traffic among EC2 instances across multiple Availability Zones but not across AWS regions

Amazon EMR Elastic Compute Cloud
Easily run and scale Apache Spark, Hive, Presto, and other big data workloads

The option that says: Use AWS Glue and store the processed data in Amazon S3 is incorrect because AWS Glue is just a serverless ETL service that crawls your data, builds a data catalog, performs data preparation, data transformation, and data ingestion. It won't allow you to utilize different big data frameworks effectively, unlike Amazon EMR. In addition, the S3 Select feature in Amazon S3 can only run simple SQL queries against a subset of data from a specific S3 object. To perform queries in the S3 bucket, you need to use Amazon Athena.

The option that says: It provides an in-memory cache that delivers up to **10x performance** improvement from milliseconds to microseconds or even at millions of requests per second is incorrect because this option **describes what Amazon DynamoDB Accelerator (DAX) does and not ElastiCache**. Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache for DynamoDB. Amazon ElastiCache cannot provide a performance improvement from milliseconds to microseconds, let alone millions of requests per second like DAX can.

Amazon S3 event notifications typically deliver events in seconds **but can sometimes take a minute or longer**. If two writes are made to a single non-versioned object at the same time, it is possible that only a single event notification will be sent. If you want to ensure that an event notification is sent for every successful write, you can enable versioning on your bucket. With versioning, every successful write will create a new version of your object and will also send an event notification.

S3 deletemarkercreated notifications are a thing. Also s3 can push publicly to MQ.

The only difference between On-Demand instances and Spot Instances is that Spot instances can be interrupted by EC2 with two minutes of notification when the EC2 needs the capacity back. On-Demand Instances let you pay for compute capacity by the hour or second (minimum of 60 seconds) with no long-term commitments. This frees you from the costs and complexities of planning, purchasing, and maintaining hardware and transforms what are commonly large fixed costs into much smaller variable costs.

S3 lifecycle has transition and expiration actions.

NACL rules are evaluated iteratively
The Rule 100 will first be evaluated. If there is a match then it will allow the request. Otherwise, it will then go to Rule 101 to repeat the same process until it goes to the default rule. In this case, when there is a request from 110.238.109.37, it will go through Rule 100 first. As Rule 100

says it will permit all traffic from any source, it will allow this request and will not further evaluate Rule 101 (which denies 110.238.109.37) nor the default rule.

There is no additional charge for using gateway endpoints. However, standard charges for data transfer and resource usage still apply. Hence, the correct answer is: Create an Amazon S3 gateway endpoint to enable a connection between the instances and Amazon S3.
The option that says: Set up a NAT Gateway in the public subnet to connect to Amazon S3 is incorrect. This will enable a connection between the private EC2 instances and Amazon S3 but it is not the most cost-efficient solution. **NAT Gateways are charged on an hourly basis even for idle time**.

Up to 15 Aurora Replicas handle read-only query traffic.Using endpoints, you can map each connection to the appropriate instance or group of instances based on your use case. For example, to perform DDL statements you can connect to whichever instance is the primary instance. To perform queries, you can connect to the reader endpoint, with Aurora automatically performing load-balancing among all the Aurora Replicas. For clusters with DB instances of different capacities or configurations, you can connect to custom endpoints associated with different subsets of DB instances. For diagnosis or tuning, you can connect to a specific instance endpoint to examine details about a specific DB instance.

A reader endpoint for an Aurora DB cluster provides load-balancing support for read-only connections to the DB cluster. Use the reader endpoint for read operations, such as queries. **By processing those statements on the read-only Aurora Replicas, this endpoint reduces the overhead on the primary instance. It also helps the cluster to scale the capacity to handle simultaneous SELECT queries, proportional to the number of Aurora Replicas in the cluster. Each Aurora DB cluster has one reader endpoint.**
I**f the cluster contains one or more Aurora Replicas, the reader endpoint load-balances each connection request among the Aurora Replicas**. In that case, you can only perform read-only statements such as SELECT in that session. If the cluster only contains a primary instance and no Aurora Replicas, the reader endpoint connects to the primary instance. In that case, you can perform write operations through the endpoint.

The option that says: Use the built-in Cluster endpoint of the Amazon Aurora database is incorrect because a **cluster endpoint (also known as a writer endpoint) simply connects to the current primary DB instance for that DB cluster.** This endpoint can perform write operations in the database such as DDL statements, which is perfect for handling production traffic but not suitable for handling queries for reporting since there will be no write database operations that will be sent.

NLB is used to distribute traffic among servers, not read replicas.

The option that says: You will be billed when your On-Demand instance is preparing to hibernate with a stopping state is correct because when the instance state is stopping, you will

not billed if it is preparing to stop however, you will still be billed if it is just preparing to hibernate.
No bill when *stopping*, but yes bill if preparing to hibernate.

Using **CloudFront Origin Access Identity** is incorrect because this is a feature which ensures that only CloudFront can serve S3 content. It does not increase throughput and ensure fast delivery of content to your customers.

Elastic Fabric Adapter
EFA brings the scalability, flexibility, and elasticity of cloud to tightly-coupled HPC applications. With EFA, tightly-coupled HPC applications have access to lower and more consistent latency and higher throughput than traditional TCP channels, enabling them to scale better. EFA support can be enabled dynamically, on-demand on any supported EC2 instance without pre-reservation, giving you the flexibility to respond to changing business/workload priorities.

AWS Site-to-Site VPN – creates an IPsec VPN connection between your VPC and your remote network. On the AWS side of the Site-to-Site VPN connection, a virtual private gateway or transit gateway provides two VPN endpoints (tunnels) for automatic failover.
AWS Client VPN – a managed client-based VPN service that provides secure TLS VPN connections between your AWS resources and on-premises networks.
AWS VPN CloudHub – capable of wiring multiple AWS Site-to-Site VPN connections together on a virtual private gateway. This is useful if you want to enable communication between different remote networks that uses a Site-to-Site VPN connection.
Third-party software VPN appliance – You can create a VPN connection to your remote network by using an Amazon EC2 instance in your VPC that's running a third party software VPN appliance.

Individual Amazon S3 objects can range in size from a minimum of 0 bytes to a maximum of 5 terabytes. The largest object that can be uploaded in a single PUT is 5 gigabytes. For objects larger than 100 megabytes, customers should consider using the Multipart Upload capability.

When setting up a bastion host in AWS, you should only allow the individual IP of the client and not the entire network. Therefore, in the Source, the proper CIDR notation should be used. The /32 denotes one IP address and the /0 refers to the entire network.

Aside from that, **network ACLs act as a firewall for your whole VPC subnet** while **security groups operate on an instance level**. Since you are securing an EC2 instance, you should be using security groups.

Cloud Formation *cfn-signal* can send success signals after *CreationPolicy* sets up everything.

EC2: not possible to enable/disable hibernation after it has been launched.

Expedited – **Expedited retrievals** allow you to quickly access your data that's stored in the S3 Glacier Flexible Retrieval storage class or the S3 Intelligent-Tiering Archive Access tier when occasional urgent requests for a subset of archives are required. For all but the largest archives (more than 250 MB), data accessed by using Expedited retrievals is typically made available within 1–5 minutes. Provisioned capacity ensures that retrieval capacity for Expedited retrievals is available when you need it. For more information, see Provisioned Capacity.
Standard – Standard retrievals allow you to access any of your archives within several hours. Standard retrievals are typically completed within 3–5 hours. This is the default option for retrieval requests that do not specify the retrieval option.
Bulk – Bulk retrievals are the lowest-cost S3 Glacier retrieval option, which you can use to retrieve large amounts, even petabytes, of data inexpensively in a day. Bulk retrievals are typically completed within 5–12 hours.
Provisioned capacity ensures that your retrieval capacity for expedited retrievals is available when you need it. Each unit of capacity provides that at least three expedited retrievals can be performed every five minutes and provides up to 150 MB/s of retrieval throughput. You should purchase provisioned retrieval capacity if your workload requires highly reliable and predictable access to a subset of your data in minutes. Without provisioned capacity Expedited retrievals are accepted, except for rare situations of unusually high demand. However, if you require access to Expedited retrievals under all circumstances, you must purchase provisioned retrieval capacity.

File Gateway presents a file-based interface to Amazon S3, which appears as a network file share. It enables you to store and retrieve Amazon S3 objects through standard file storage protocols. File Gateway allows your existing file-based applications or devices to use secure and durable cloud storage without needing to be modified. With File Gateway, your configured S3 buckets will be available as Network File System (NFS) mount points or Server Message Block (SMB) file shares.
The option that says: Use the AWS **Storage Gateway** volume gateway to store the backup data and directly access it using Amazon S3 API actions is incorrect. Although this is a possible solution, you cannot directly access the volume gateway using Amazon S3 APIs. You should use File Gateway to access your data in Amazon S3.

File Gateway is on client end.
Storage end is on cloud end.

Magnetic volumes are ideal for workloads where data are accessed infrequently, and applications where the lowest storage cost is important. Take note that this is a Previous Generation Volume. The latest low-cost magnetic storage types are C**old HDD (sc1) and Throughput Optimized HDD (st1)** volumes.

SWF is incorrect because this is a fully-managed state tracker and task coordinator service. It does not provide serverless orchestration to multiple AWS resources.

AWS **Step Functions** provides serverless orchestration for modern applications. Orchestration centrally manages a workflow by breaking it into multiple steps, adding flow logic, and tracking the inputs and outputs between the steps. As your applications execute, Step Functions maintains application state, tracking exactly which workflow step your application is in, and stores an event log of data that is passed between application components. That means that if networks fail or components hang, your application can pick up right where it left off. Application development is faster and more intuitive with Step Functions, because you can define and manage the workflow of your application independently from its business logic. Making changes to one does not affect the other. You can easily update and modify workflows in one place, without having to struggle with managing, monitoring and maintaining multiple point-to-point integrations. Step Functions frees your functions and containers from excess code, so your applications are faster to write, more resilient, and easier to maintain.

HPC not supported on Windows
The OS-bypass **capabilities of EFAs are not supported on Windows instances. If you attach an EFA to a Windows instance, the instance functions as an Elastic Network Adapter, without the added EFA capabilities**.
Elastic Network Adapters (ENAs) provide traditional IP networking features that are required to support VPC networking. EFAs provide all of the same traditional IP networking features as ENAs, and they also support OS-bypass capabilities. OS-bypass enables HPC and machine learning applications to bypass the operating system kernel and to communicate directly with the EFA device.

To ensure that your users access your files using only CloudFront URLs, regardless of whether the URLs are signed, use **Origin Access Identity**.

Multipart Upload is incorrect because this feature simply allows you to upload a single object as a set of parts. You can upload these object parts independently and in any order. If transmission of any part fails, you can retransmit that part without affecting other parts. After all parts of your object are uploaded, Amazon S3 assembles these parts and creates the object. In general, when your object size reaches 100 MB, you should consider using multipart uploads instead of uploading the object in a single operation.

AWS Site-to-Site VPN – creates an IPsec VPN connection between your VPC and your remote network. On the AWS side of the Site-to-Site VPN connection, a virtual private gateway or transit gateway provides two VPN endpoints (tunnels) for automatic failover.

AWS VPN CloudHub – **capable of wiring multiple AWS Site-to-Site VPN connections together on a virtual private gateway**. This is useful if you want to enable communication between different remote networks that uses a Site-to-Site VPN connection.

When you launch an instance in a VPC, **you can assign up to five security groups to the instance**. Security groups act at the instance level, not the subnet level. Therefore, each instance in a subnet in your VPC can be assigned to a different set of security groups.

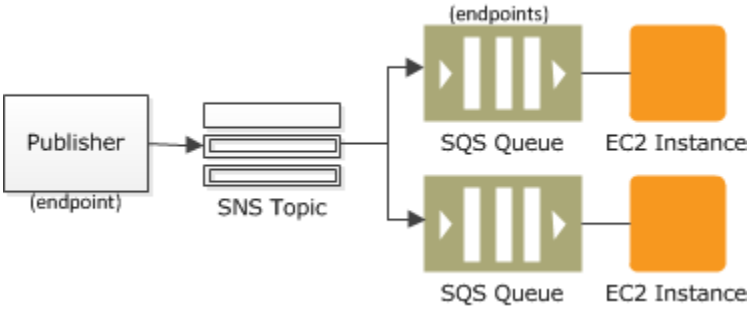
AnyCast IP address is primarily used for AWS Global Accelerator and not for security group configurations.

Take note that the **Network ACL covers the entire subnet** which means that other applications that use the same subnet will also be affected.

Provisioned IOPS (SSD) volumes offer storage with consistent and low-latency performance and are designed for I/O intensive applications such as large relational or NoSQL databases. Magnetic volumes provide the lowest cost per gigabyte of all EBS volume types.

Magnetic volumes are ideal for workloads where data are accessed infrequently, and applications where the lowest storage cost is important. Take note that this is a Previous Generation Volume. The latest low-cost magnetic storage types are Cold HDD (sc1) and Throughput Optimized HDD (st1) volumes.

Configure AWS Config to trigger an evaluation that will check the compliance for a user's password periodically.



SNS does pub-sub for multiple SQS queues (fan out method). After the visibility timer is over the message becomes available to be queued again unless it was explicitly deleted after being processed.

Amazon SQS supports *dead-letter queues* (DLQ), which other queues (*source queues*) can target for messages that can't be processed (consumed) successfully. Dead-letter queues are useful for debugging your application or messaging system because they let you isolate unconsumed messages to determine why their processing doesn't succeed.

S3 DynamoDB -> **Gateway** endpoints.

Rest use interface endpoints.

To **avoid the NAT Gateway Data Processing charge**, you could **set up a Gateway Type VPC endpoint** and route the traffic to/from S3 through the VPC endpoint instead of going through the NAT Gateway.

There is no data processing or hourly charges for using Gateway Type VPC endpoints.

You can use Run Command from the console to configure instances without having to login to each instance.

By using Cached volumes, you store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally in your on-premises network. Cached volumes offer substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data. This is the best solution for this scenario.

VPC endpoints are region specific.

Transit Gateway is used for **interconnecting VPCs and on-premises networks** through a central hub

Since **security groups are stateful**, you can apply any changes to an incoming rule and it will be automatically applied to the outgoing rule. **What goes in will go out.**

AWS Backup -> 90 days

Aurora -> RPO 1 second RTO 1 minute
RTO is time it takes to get back up
RPO is how much data you can lose

32768-61000 for a general safe outbound port range (or 1024-65535).

Amazon Data Lifecycle Manager -> automatically creates EBS snapshots

Amazon Detective automatically collects log data from your AWS resources to analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities in your AWS account. A firewall must be created at the VPC level and not at the subnet level.
Q. A business plans to deploy an application on EC2 instances within an Amazon VPC and is considering adopting a Network Load Balancer to distribute incoming traffic among the instances. A solutions architect needs to suggest a solution that will enable the security team to inspect traffic entering and exiting their VPC.

Ans: Create a firewall using the AWS Network Firewall service at the VPC level then add custom rule groups for inspecting ingress and egress traffic. Update the necessary VPC route tables.

An internet gateway serves two purposes: to provide a target in your VPC route tables for internet-routable traffic, and to perform network address translation (NAT) for instances that have been assigned public IPv4 addresses.

Cross Zone Load Balancing -> balances across ALL available units

To summarize the difference in throughput vs. IOPS**, IOPS is a count of the read/write operations** per second, but **throughput is the actual measurement of read/write bits per second that are transferred** over a network.

SAML 2.0 is a standard that is used mostly for on-premise systems, usually Microsoft Active Directory or others, so in this case users can log into AWS with their on-premise credentials. Web Identity Federation is where we use an IDP (like Amazon, Google, etc.)

In this scenario, the best option is to group the set of users in an IAM Group and then apply a policy with the required access to the Amazon S3 bucket. This will enable you to easily add, remove, and manage the users instead of manually adding a policy to each and every 100 IAM users.
IAM groups can have policies (read: actual permissions) attached to them. IAM roles are more specific.

With geoproximity routing you can specify a coverage area, you can't do that with geolocation routing (that applies to a whole country)

kinesis 24 hours

Amazon S3 Access Points

ebs non-root persists?

sqs and swf decouple applications, not rds/databases

Network ACL is for entire VPC, Security Group is for individual services

Data Firehose can push to S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk

ELB only runs in one region

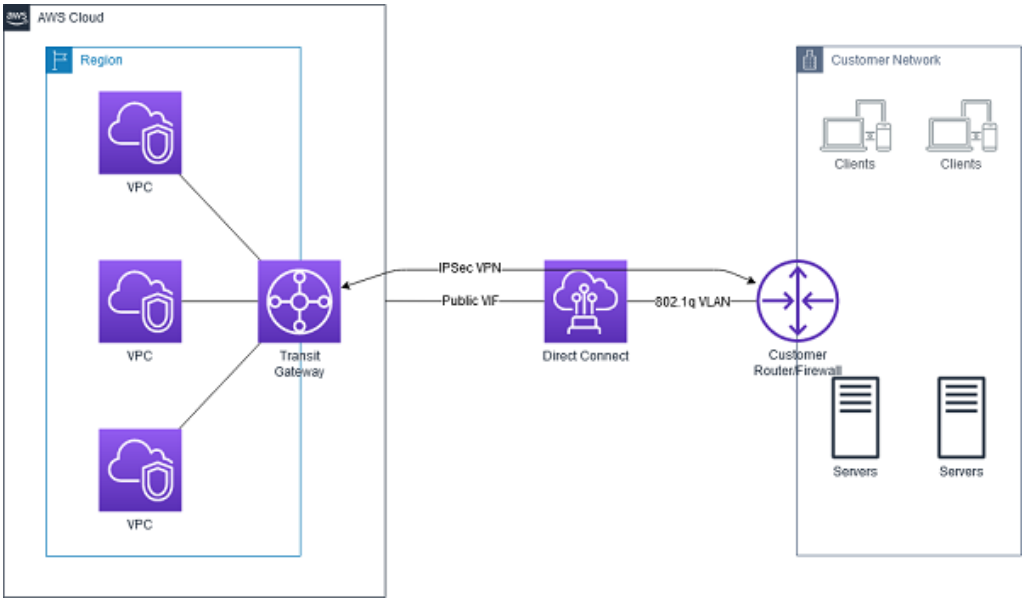
You cannot set up an Active-Active Failover with One Primary and One Secondary Resource. Remember that an Active-Active Failover uses all available resources all the time without a primary nor a secondary resource.

ACTIVE ACTIVE HAS TO BE WEIGHTED CAN'T BE PRIMARY SECONDARY

mysql rds allows for autoscaling as well (don't manually increase)

Amazon FSx for Windows File Server is incorrect. This won't provide low-latency access since all the files are stored on AWS, which means that they will be accessed via the internet. **AWS Storage Gateway supports local caching** without any development overhead making it suitable for low-latency applications.

Enabling In-Memory Acceleration with DynamoDB Accelerator (DAX) is incorrect because the DAX feature is primarily used for read performance improvement of your DynamoDB table from milliseconds response time to microseconds. It does not have any relationship with Amazon Kinesis Data Stream in this scenario. (Hence, increasing the write capacity assigned to the shard table is the correct answer.)



FINISH