

# **AWS SAA – C02 Exam Questions**

**Ajwad Javed**

<https://github.com/ajwadjaved>

## Stephane Exam

CloudFront uses Edge Locations to cache content while Global Accelerator uses Edge Locations to find an optimal pathway to the nearest regional endpoint. CloudFront is designed to handle HTTP protocol meanwhile **Global Accelerator is best used for both HTTP and non-HTTP protocols such as TCP and UDP.**

Ad hoc queries -> single results.

Amazon EMR -> run big data stuff

Cloudhub -> communicate safely

Cognito

User pools are for authentication (identity verification). With a user pool, your app users can sign in through the user pool or federate through a third-party identity provider (IdP).

Identity pools are for authorization (access control). You can use identity pools to create unique identities for users and give them access to other AWS services.

Use Amazon Cognito Identity Pools - The two main components of Amazon Cognito are user pools and identity pools. **Identity pools provide AWS credentials to grant your users access to other AWS services.** To enable users in your user pool to access AWS resources, you can configure an identity pool to exchange user pool tokens for AWS credentials. So, identity pools aren't an authentication mechanism in themselves and hence aren't a choice for this use case.

AWS GuardDuty doesn't scan CloudWatch. Does do **CloudTrail, VPC Flow Logs, DNS Logs.**

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 in AWS.

Gateway endpoints: s3, dynamoDB

VPC sharing -> shares subnets not VPCs

Sns pub/sub, more emails to subscribed topics

dynamoDB noSQL while aurora is relational

Lambda layers provide a convenient way to package libraries and other dependencies that you can use with your Lambda functions. Using layers reduces the size of uploaded deployment archives and makes it faster to deploy your code. A layer is a .zip file archive that can contain additional code or data.

Type of Access Control	AWS Account Level Control	User Level Control
IAM Policies	No	Yes
ACLs	Yes	No
Bucket Policies	Yes	Yes

By default, FIFO queues support up to **3,000 messages** per second with batching, or up to **300 messages** per second (300 send, receive, or delete operations per second) without batching. Therefore, using batching you can meet a throughput requirement of upto 3,000 messages per second.

### Design High Performing Architecture

CloudFront supports HTTP/Messaging.  
Global Accelerator is for gaming, IoT, UDP, and the like.

By default, the **root volume for an AMI backed by Amazon EBS is deleted** when the **instance terminates**. You can **change the default behavior to ensure that the volume persists** after the instance terminates. **Non-root EBS volumes remain available** even after you terminate an instance to which the volumes were attached. Therefore, this option is correct. Instance Store however terminates along with the EC2 volume.

On termination of an EC2 instance, all the attached EBS volumes are always terminated - As mentioned earlier, non-root EBS volumes remain available even after you terminate an instance to which the volumes were attached. Hence this option is incorrect.

**FSx for Windows does not allow you to present S3 objects as files and does not allow you to write changed data back to S3. Therefore you cannot reference the "cold data" with quick access for reads and updates at low cost.**

Currently, Amazon S3 can publish notifications for the following events: New object-created events, Object removal events, Restore object events, Reduced Redundancy Storage (RRS) object lost events, Replication events.

Amazon S3 supports the following destinations where it can publish events: Amazon Simple Notification Service (Amazon SNS), Amazon Simple Queue Service (Amazon SQS) queue, AWS Lambda

User Data is generally used to perform common automated configuration tasks and even run scripts after the instance starts. When you launch an instance in Amazon EC2, you can pass two types of user data - shell scripts and cloud-init directives. You can also pass this data into the launch wizard as plain text or as a file.

**By default, scripts entered as user data are executed with root user privileges -**

Scripts entered as user data are executed as the root user, hence do not need the sudo command in the script. Any files you create will be owned by root; if you need non-root users to have file access, you should modify the permissions accordingly in the script.

**By default, user data runs only during the boot cycle when you first launch an instance -** By default, user data scripts and cloud-init directives run only during the boot cycle when you first launch an instance. You can update your configuration to ensure that your user data scripts and cloud-init directives run every time you restart your instance.

Incorrect options:

**By default, user data is executed every time an EC2 instance is re-started -** As discussed above, this is not a default configuration of the system. But, can be achieved by explicitly configuring the instance.

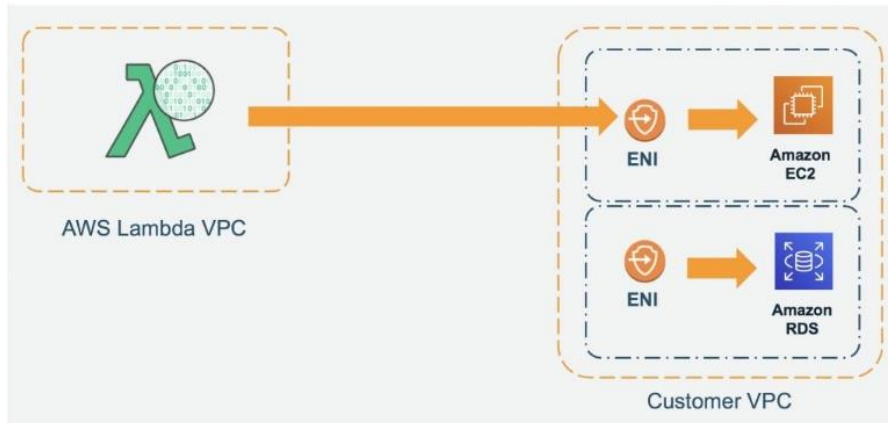
**When an instance is running, you can update user data by using root user credentials -** You can't change the user data if the instance is running (even by using root user credentials), but you can view it.

**By default, scripts entered as user data do not have root user privileges for executing -** Scripts entered as user data are executed as the root user, hence do not need the sudo command in the script.

Can't edit user data when an instance is running even with root. Only view.

## Tip #1: When to VPC-Enable a Lambda Function

Lambda functions always operate from an AWS-owned VPC. By default, your function has full ability to make network requests to any public internet address — this includes access to any of the public AWS APIs. For example, your function can interact with AWS DynamoDB APIs to PutItem or Query for records. You should only enable your functions for VPC access when you need to interact with a private resource located in a private subnet. An RDS instance is a good example.



Once your function is VPC-enabled, all network traffic from your function is subject to the routing rules of your VPC/Subnet. If your function needs to interact with a public resource, you will need a route through a NAT gateway in a public subnet.

You can configure your Lambda function to pull in additional code and content in the form of layers. A layer is a ZIP archive that contains libraries, a custom runtime, or other dependencies. With layers, you can use libraries in your function without needing to include them in your deployment package. Layers let you keep your deployment package small, which makes development easier. A function can use up to 5 layers at a time.

NAT instance can be used as a bastion server  
Security Groups can be associated with a NAT instance  
NAT instance supports port forwarding

Kinesis Data Firehose can only write to S3, Redshift, Elasticsearch or Splunk. You can't have applications consuming data streams from Kinesis Data Firehose, that's the job of Kinesis Data Streams.

**Q: When should I use Amazon Kinesis Data Streams, and when should I use Amazon SQS?**

We recommend Amazon Kinesis Data Streams for use cases with requirements that are similar to the following:

- Routing related records to the same record processor (as in streaming MapReduce). For example, counting and aggregation are simpler when all records for a given key are routed to the same record processor.
- Ordering of records. For example, you want to transfer log data from the application host to the processing/archival host while maintaining the order of log statements.
- Ability for multiple applications to consume the same stream concurrently. For example, you have one application that updates a real-time dashboard and another that archives data to Amazon Redshift. You want both applications to consume data from the same stream concurrently and independently.
- Ability to consume records in the same order a few hours later. For example, you have a billing application and an audit application that runs a few hours behind the billing application. Because Amazon Kinesis Data Streams stores data for up to 7 days, you can run the audit application up to 7 days behind the billing application.

We recommend Amazon SQS for use cases with requirements that are similar to the following:

- Messaging semantics (such as message-level ack/fail) and visibility timeout. For example, you have a queue of work items and want to track the successful completion of each item independently. Amazon SQS tracks the ack/fail, so the application does not have to maintain a persistent checkpoint/cursor. Amazon SQS will delete acked messages and redeliver failed messages after a configured visibility timeout.
- Individual message delay. For example, you have a job queue and need to schedule individual jobs with a delay. With Amazon SQS, you can configure individual messages to have a delay of up to 15 minutes.
- Dynamically increasing concurrency/throughput at read time. For example, you have a work queue and want to add more readers until the backlog is cleared. With Amazon Kinesis Data Streams, you can scale up to a sufficient number of shards (note, however, that you'll need to provision enough shards ahead of time).
- Leveraging Amazon SQS's ability to scale transparently. For example, you buffer requests and the load changes as a result of occasional load spikes or the natural growth of your business. Because each buffered request can be processed independently, Amazon SQS can scale transparently to handle the load without any provisioning instructions from you.

**VPC Flow Logs:** VPC Flow Logs is a feature that enables you to capture information about the IP traffic going to and from network interfaces in your VPC. Flow log data is used to analyze network traces and helps with network security. Flow log data can be published to Amazon CloudWatch Logs or Amazon S3. You cannot use VPC Flow Logs to debug and trace data across accounts.

**CloudTrail:** With CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure. You can use AWS CloudTrail to answer questions such as - "Who made an API call to modify this resource?". CloudTrail provides event history of your AWS account activity thereby enabling governance, compliance, operational auditing, and risk auditing of your AWS account. You cannot use CloudTrail to debug and trace data across accounts.

S3 cannot directly write data into SNS, although it can certainly use S3 event notifications to send an event to SNS. Also, **SNS cannot directly send messages to Kinesis Data Streams** (it can to **Firehose**). So this option is incorrect.

—

## Design Resilient Architecture

The Amazon Redshift cluster must be in the same AWS account and the same AWS Region as the replication instance. During a database migration to Amazon Redshift, AWS DMS first moves data to an Amazon S3 bucket. When the files reside in an Amazon S3 bucket, AWS DMS then transfers them to the proper tables in the Amazon Redshift data warehouse. AWS DMS creates the S3 bucket in the same AWS Region as the Amazon Redshift database. The AWS DMS replication instance must be located in that same region.

### Kinesis Data Stream Limitations

However, the user is expected to manually provision an appropriate number of shards to process the expected volume of the incoming data stream. The throughput of an Amazon Kinesis data stream is designed to scale without limits via increasing the number of shards within a data stream. Therefore Kinesis Data Streams is not the right fit for this use-case.

Use EC2 user data to install the application at boot time - User data of an instance can be used to perform common automated configuration tasks or run scripts after the instance starts. User data, cannot, however, be used to install the application since it takes over 45 minutes for the installation which contains static as well as dynamic files that must be generated during the installation process.

Aurora Replicas have two main purposes. You can issue queries to them to scale the read operations for your application. You typically do so by connecting to the reader endpoint of the cluster. That way, Aurora can spread the load for read-only connections across as many Aurora Replicas as you have in the cluster. Aurora Replicas also help to increase availability. If the writer instance in a cluster becomes unavailable, Aurora automatically promotes one of the reader instances to take its place as the new writer. **Up to 15 Aurora Replicas can be distributed across the Availability Zones that a DB cluster spans within an AWS Region.**

**API Gateway, SQS, Kinesis** can be used to *throttle* requests.

Amazon API Gateway, Amazon SQS and Amazon Kinesis - To prevent your API from being overwhelmed by too many requests, Amazon API Gateway throttles requests to your API using the token bucket algorithm, where a token counts for a request. Specifically, API Gateway sets a limit on a steady-state rate and a burst of request submissions against all APIs in your account. In the token bucket algorithm, the burst is the maximum bucket size.

Amazon SQS - Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. Amazon SQS offers buffer capabilities to smooth out temporary volume spikes without losing messages or increasing latency.

Amazon Kinesis - Amazon Kinesis is a fully managed, scalable service that can ingest, buffer, and process streaming data in real-time.

Activate read-through caching on the Amazon Aurora database - **Aurora does not have built-in support for read-through caching**, so this option just serves as a distractor. To implement caching, you will need to integrate something like ElastiCache and that would need code changes for the application.

Convert the existing standard queue into a FIFO queue - You can't convert an existing standard queue into a FIFO queue.

Global Accelerator can do weighted traffic routing for blue/green deployment.

It provides two static anycast IP addresses that act as a fixed entry point to your application endpoints in a single or multiple AWS Regions, such as your Application Load Balancers, Network Load Balancers, Elastic IP addresses or Amazon EC2 instances, in a single or in multiple AWS regions.

**Route53 would use cached DNS for a long time so not suitable for blue/green deployment.**

**ELB wouldn't work if you need it as a global solution**

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.**

S3 replication only supports copying new Amazon S3 objects after it is enabled. Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS account or by different accounts. Object may be replicated to a single destination bucket or multiple destination buckets. Destination buckets can be in different AWS Regions or within the same Region as the source bucket.

Auto Scaling group lifecycle hooks enable you to perform custom actions as the Auto Scaling group launches or terminates instances. Lifecycle hooks enable you to perform custom actions by pausing instances as an Auto Scaling group launches or terminates them. When an instance is paused, it remains in a wait state either until you complete the lifecycle action using the complete-lifecycle-action command or the CompleteLifecycleAction operation, or until the timeout period ends (one hour by default). For example, you could install or configure software on newly launched instances, or download log files from an instance before it terminates. *Scheduled action is only for scaling actions based on some schedule. Not for custom actions.*

Spread placement group: **SEVEN** instances per AZ.

SQS can't send notification emails, SNS can.



Amazon S3 delivers strong read-after-write consistency automatically, without changes to performance or availability, without sacrificing regional isolation for applications, and at no additional cost.

## NO DEFAULT CACHE.

—

## Designing Secure Applications and Architectures

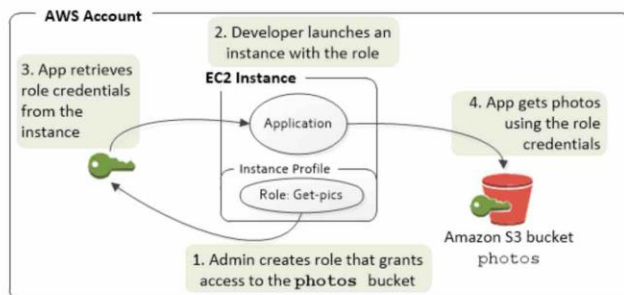
VPN tunnel: An encrypted link where data can pass from the customer network to or from AWS.

[Manage IAM roles](#) - access to cross account resources with IAM roles is allowed,

Temporary Access to complete the task.

### How Do Roles for EC2 Instances Work?

In the following figure, a developer runs an application on an EC2 instance that requires access to the S3 bucket named photos. An administrator creates the Get-pics service role and attaches the role to the EC2 instance. The role includes a permissions policy that grants read-only access to the specified S3 bucket. It also includes a trust policy that allows the EC2 instance to assume the role and retrieve the temporary credentials. When the application runs on the instance, it can use the role's temporary credentials to access the photos bucket. The administrator doesn't have to grant the developer permission to access the photos bucket, and the developer never has to share or manage credentials.



You can use two types of VPC endpoints to access Amazon S3: gateway endpoints and interface endpoints. A gateway endpoint is a gateway that you specify in your route table to access Amazon S3 from your VPC over the AWS network. Interface endpoints extend the functionality of gateway endpoints by using private IP addresses to route requests to Amazon S3 from within your VPC, on premises, or from a VPC in another AWS Region using VPC peering or AWS Transit Gateway.

You must remember that these two services (**S3, DynamoDB**) use a **VPC gateway** endpoint. The rest of the AWS services use VPC interface endpoints.

As the existing infrastructure is within AWS Cloud, therefore a VPN connection is not required.

VPC sharing (part of Resource Access Manager) allows **multiple AWS accounts** to create their application resources such as EC2 instances, RDS databases, Redshift clusters, and Lambda

functions, into shared and centrally-managed Amazon Virtual Private Clouds (VPCs). To set this up, the account that owns the VPC (owner) shares one or more subnets with other accounts (participants) that belong to the same organization from AWS Organizations. After a subnet is shared, the participants can view, create, modify, and delete their application resources in the subnets shared with them. Participants cannot view, modify, or delete resources that belong to other participants or the VPC owner.

You can share Amazon VPCs to leverage the implicit routing within a VPC for applications that require a high degree of interconnectivity and are within the same trust boundaries. This reduces the number of VPCs that you create and manage while using separate accounts for billing and access control.

#### Permission Boundaries (basic)

A permissions boundary can be used to control the maximum permissions employees can grant to the IAM principals (that is, users and roles) that they create and manage. As the IAM administrator, you can define one or more permissions boundaries using managed policies and allow your employee to create a principal with this boundary. The employee can then attach a permissions policy to this principal. However, the effective permissions of the principal are the intersection of the permissions boundary and permissions policy. As a result, the new principal cannot exceed the boundary that you defined. Therefore, using the permissions boundary offers the right solution for this use-case.

VPC peering facilitates a connection between **two VPCs within the AWS network**, therefore this option cannot be used to send and receive data between the remote branch offices of the company

Use Amazon Cognito Identity Pools - The two main components of Amazon Cognito are user pools and identity pools. Identity pools provide AWS credentials to grant your users access to other AWS services. To enable users in your user pool to access AWS resources, you can configure an identity pool to exchange user pool tokens for AWS credentials. So, identity pools aren't an authentication mechanism in themselves and hence aren't a choice for this use case.

#### **IAM Auth does not support ElastiCache.**

ElastiCache Memcached cannot be used as a cache to serve static content from S3, so both these options are incorrect.

Use CloudFront to improve performance.

By using Amazon S3 Analytics Storage Class analysis you can analyze storage access patterns to help you decide when to transition the right data to the right storage class. This new Amazon S3 analytics feature observes data access patterns to help you determine when to transition less frequently accessed STANDARD storage to the STANDARD\_IA (IA, for infrequent access) storage class. **Storage class analysis does not give recommendations for transitions to the ONEZONE\_IA or S3 Glacier storage classes.** (cheaper options)

Better option: *Use AWS Cost Explorer Resource Optimization to get a report of EC2 instances that are either idle or have low utilization and use AWS Compute Optimizer to look at instance type recommendations.*

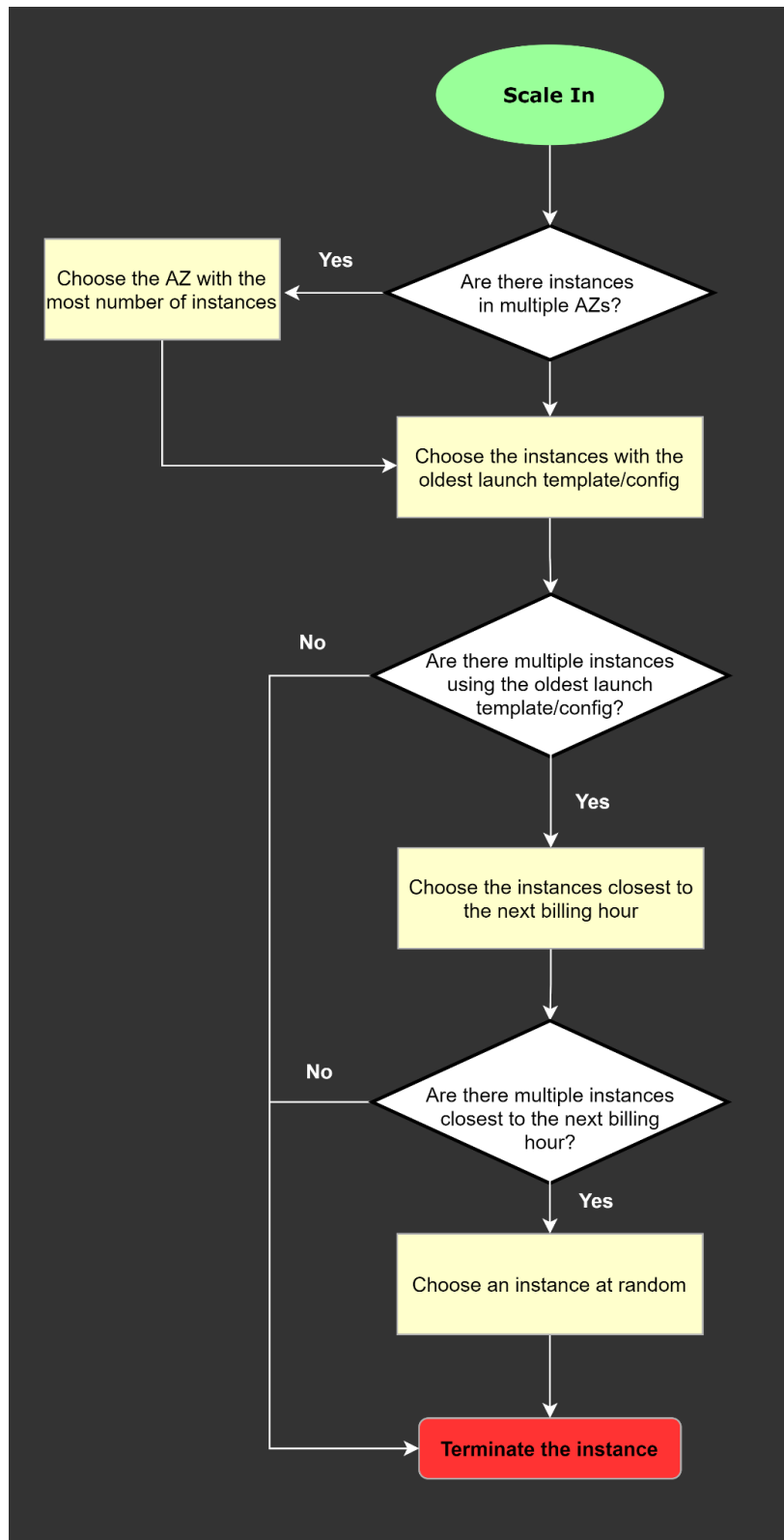
A launch template is similar to a launch configuration, in that it specifies instance configuration information such as the ID of the Amazon Machine Image (AMI), the instance type, a key pair, security groups, and the other parameters that you use to launch EC2 instances. Also, defining a launch template instead of a launch configuration allows you to have multiple versions of a template.

**With launch templates (not launch configuration), you can provision capacity** across multiple instance types using both On-Demand Instances and Spot Instances to achieve the desired scale, performance, and cost. Hence this is the correct option.

---

## Tutorial Dojo Exam 1

### EC2 Termination Flowchart



The option that says: *Enable the Enhanced Monitoring option in EC2 and install CloudWatch agent to all the EC2 instances* to be able to view the memory and disk utilization in the

CloudWatch dashboard is incorrect because **Enhanced Monitoring is a feature of Amazon RDS**. By default, Enhanced Monitoring metrics are stored for 30 days in the CloudWatch Logs.

The option that says: *Use Amazon Inspector and install the Inspector agent* to all EC2 instances is incorrect because **Amazon Inspector is an automated security assessment service that helps you test the network accessibility** of your Amazon EC2 instances and the security state of your applications running on the instances. It does not provide a custom metric to track the memory and disk utilization of each and every EC2 instance in your VPC.

AWS WAF is tightly integrated with Amazon CloudFront, the Application Load Balancer (ALB), Amazon API Gateway, and AWS AppSync – services that AWS customers commonly use to deliver content for their websites and applications. When you use AWS WAF on Amazon CloudFront, your rules run in all AWS Edge Locations, located around the world close to your end-users. This means security doesn't come at the expense of performance. Blocked requests are stopped before they reach your web servers. When you use AWS WAF on regional services, such as Application Load Balancer, Amazon API Gateway, and AWS AppSync, your rules run in the region and can be used to protect Internet-facing resources as well as internal resources.

The option that says: Create a custom network ACL and associate it with the subnet of the Application Load Balancer to block the offending requests is incorrect. Although NACLs can help you block incoming traffic, this option wouldn't be able to limit the number of requests from a single IP address that is dynamically changing.

The option that says: Create a custom rule in the security group of the Application Load Balancer to block the offending requests is incorrect because the security group can only allow incoming traffic. Remember that you can't deny traffic using security groups. In addition, it is not capable of limiting the rate of traffic to your application unlike AWS WAF.

To control the traffic coming in and out of your VPC network, you can use the network access control list (ACL). It is an optional layer of security for your VPC that acts as a firewall for controlling traffic in and out of one or more subnets. This is the best solution among other options as you can easily add and remove the restriction in a matter of minutes.

Although the use of Secrets Manager in securing sensitive data in ECS is valid, **Amazon ECS doesn't support resource-based policies**. An example of a resource-based policy is the S3 bucket policy. An **ECS task assumes an execution role (IAM role)** to be able to call other AWS services like AWS Secrets Manager on your behalf.

Take note that to achieve fault tolerance, you need to have redundant resources in place to avoid any system degradation in the event of a server fault or an Availability Zone outage. Having a fault-tolerant architecture entails an extra cost in running additional resources than what is usually needed. This is to ensure that the mission-critical workloads are processed.

RDS Read Replica is incorrect as a Read Replica provides an asynchronous replication instead of synchronous.

## Multi-AZ Deployments

Synchronous replication – highly durable

Only database engine on primary instance is active

Automated backups are taken from standby

Always span two Availability Zones within a single Region

Database engine version upgrades happen on primary

Automatic failover to standby when a problem is detected

In the scenario, you can trigger a Lambda function whenever a listing is deleted from the database. You can then write the logic of the function to send the listing data to an SQS queue and have different processes consume it.

Hence, the correct answer is: Create a native function or a stored procedure that invokes a Lambda function. Configure the Lambda function to send event notifications to an Amazon SQS queue for the processing system to consume.

RDS events only provide operational events such as DB instance events, DB parameter group events, DB security group events, and DB snapshot events. What we need in the scenario is to capture data-modifying events (INSERT, DELETE, UPDATE) which can be achieved thru native functions or stored procedures.

Remember S3 lifecycle policy hierarchy. Has to go through IA for glacier.

CloudFront signed URLs and signed cookies provide the same basic functionality: they allow you to control who can access your content. If you want to serve private content through CloudFront and you're trying to decide whether to use signed URLs or signed cookies, consider the following:

Use signed URLs for the following cases:

- You want to use an RTMP distribution. **Signed cookies aren't supported for RTMP distributions.**
- You want to **restrict access to individual files**, for example, an installation download for your application.
- Your users are using a client (for example, a custom HTTP client) that doesn't support cookies.

Use signed cookies for the following cases:



- You want to **provide access to multiple restricted files**, for example, all of the files for a video in HLS format or all of the files in the subscribers' area of a website.
- You don't **want to change your current URLs.**

**Amazon Aurora Serverless** is an on-demand, auto-scaling configuration for Amazon Aurora. An Aurora Serverless DB cluster is a DB cluster that automatically starts up, shuts down, and scales up or down its compute capacity based on your application's needs. Aurora Serverless provides a relatively simple, cost-effective option for infrequent, intermittent, sporadic or unpredictable workloads. It can provide this because it automatically starts up, scales compute capacity to match your application's usage and shuts down when it's not in use.

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

The option that says: Launch an Amazon Aurora Provisioned DB cluster with burstable performance DB instance class types is incorrect because an Aurora Provisioned DB cluster is not suitable for intermittent, sporadic, and unpredictable transactional workloads. This model works well when the database workload is predictable because you can adjust capacity manually based on the expected workload. A better database setup here is to use an Amazon Aurora Serverless cluster.

The option that says: Upload the data to the closest S3 bucket. Set up a cross-region replication and copy the objects to the destination bucket is incorrect because **replicating the objects to the destination bucket takes about 15 minutes**. Take note that the requirement in the scenario is to aggregate the data in the fastest way.

FEATURES	SSD Solid State Drive	HDD Hard Disk Drive
Best for workloads with:	<i>small, random</i> I/O operations	<i>large, sequential</i> I/O operations
Can be used as a bootable volume?	Yes	No
Suitable Use Cases	<ul style="list-style-type: none"> <li>- Best for <b>transactional workloads</b></li> <li>- Critical business applications that require sustained IOPS performance</li> <li>- Large database workloads such as MongoDB, Oracle, Microsoft SQL Server and many others...</li> </ul>	<ul style="list-style-type: none"> <li>- Best for <i>large streaming workloads</i> requiring consistent, fast throughput at a low price</li> <li>- Big data, Data warehouses, Log processing</li> <li>- Throughput-oriented storage for large volumes of data that is <i>infrequently</i> accessed</li> </ul>
Cost	moderate / high 	low 
Dominant Performance Attribute	IOPS	Throughput (MiB/s)

Considering that the company is using a corporate Active Directory, it is best to use **AWS Directory Service AD Connector** for easier integration. In addition, since the roles are already assigned using groups in the corporate Active Directory, it would be better to also use IAM Roles. Take note that you can assign an IAM Role to the users or groups from your Active Directory once it is integrated with your VPC via the AWS Directory Service AD Connector. AWS Directory Service Simple AD is incorrect because this just provides a subset of the features offered by AWS Managed Microsoft AD, including the ability to manage user accounts and group memberships, create and apply group policies, securely connect to Amazon EC2 instances, and provide Kerberos-based single sign-on (SSO). In this scenario, the more suitable component to use is the AD Connector since it is a directory gateway with which you can redirect directory requests to your on-premises Microsoft Active Directory. IAM Groups is incorrect because this is just a collection of IAM users. Groups let you specify permissions for multiple users, which can make it easier to manage the permissions for those users. In this scenario, the more suitable one to use is IAM Roles in order for permissions to create AWS Directory Service resources.

### DIRECT CONNCTET IS FOR ON-PREMISES TO CLOUD

The option that says: Use AWS Transit Gateway to route all access in S3 and DynamoDB to a public endpoint is incorrect because a Transit Gateway simply connects your VPC and on-



premises networks through a central hub. It acts as a cloud router that allows you to integrate multiple networks.

The option that says: Use AWS Direct Connect to route all access to S3 and DynamoDB via private endpoints is incorrect because **AWS Direct Connect is primarily used to establish a dedicated network connection from your premises to AWS**. The scenario didn't say that the company is using its on-premises server or has a hybrid cloud architecture.

The option that says: Use AWS VPN CloudHub to route all access in S3 and DynamoDB to a private endpoint is incorrect because **AWS VPN CloudHub is mainly used to provide secure communication between remote sites** and not for creating a private endpoint to access Amazon S3 and DynamoDB within the Amazon network.

A **file gateway** supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. By using this combination, you can store and retrieve objects in Amazon S3 using industry-standard file protocols such as Network File System (NFS) and Server Message Block (SMB). The software appliance, or gateway, is deployed into your on-premises environment as a virtual machine (VM) running on VMware ESXi, Microsoft Hyper-V, or Linux Kernel-based Virtual Machine (KVM) hypervisor.

Always remember that the messages in the SQS queue will continue to exist even after the EC2 instance has processed it, until you delete that message. You have to ensure that you delete the message after processing to prevent the message from being received and processed again once the visibility timeout expires.

There are three main parts in a distributed messaging system:

1. The components of your distributed system (EC2 instances)
2. Your queue (distributed on Amazon SQS servers)
3. Messages in the queue.

DynamoDB

**Using partition keys with low-cardinality attributes**, which have a few number of distinct values for each item is incorrect because this is the exact opposite of the correct answer.

**Remember that the more distinct partition key values your workload accesses, the more those requests will be spread across the partitioned space.** Conversely, the less distinct partition key values, the less evenly spread it would be across the partitioned space, which effectively slows the performance.

---

## Tutorial Dojo Exam 2

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.

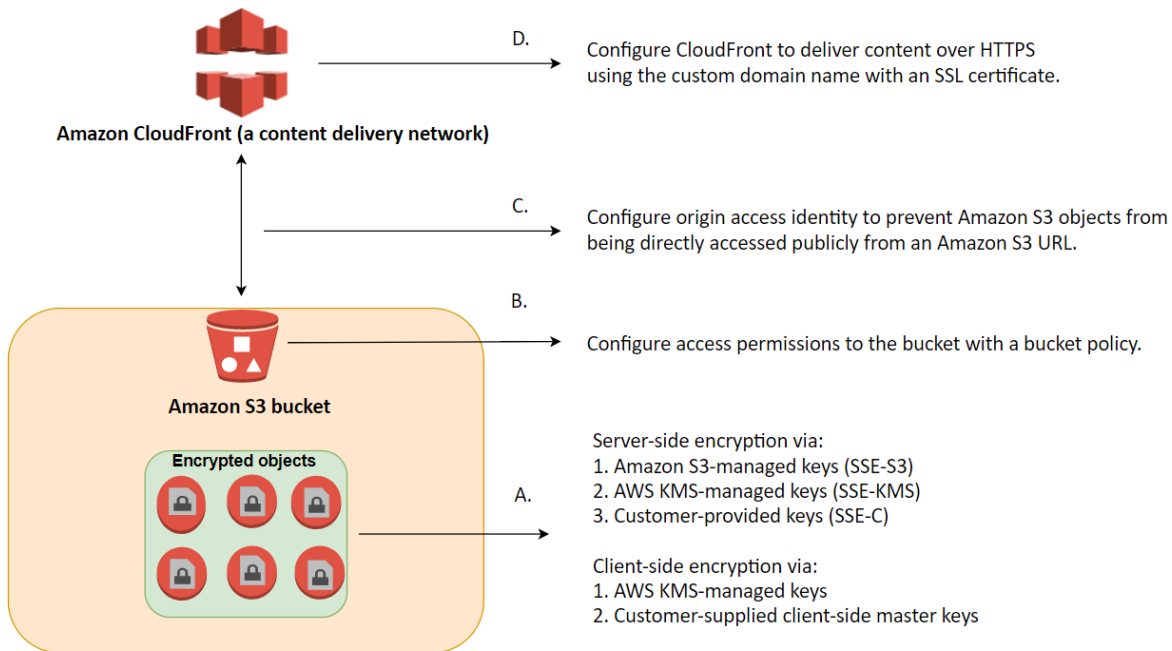
A single Amazon SQS message queue can contain an unlimited number of messages. However, there is a 120,000 limit for the number of inflight messages for a standard queue and 20,000 for a FIFO queue. Messages are inflight after they have been received from the queue by a consuming component, but have not yet been deleted from the queue.

DynamoDB increase write capacity assigned to the shard table.

If the shard iterator expires immediately before you can use it, this might indicate that the DynamoDB table used by Kinesis does not have enough capacity to store the lease data. This situation is more likely to happen if you have a large number of shards. To solve this problem, increase the write capacity assigned to the shard table.

Using **AWS Firewall Manager** to set up security rules that block SQL injection and cross-site scripting attacks, then associating the rules to the Application Load Balancer is incorrect because AWS Firewall Manager just simplifies your AWS WAF and AWS Shield Advanced administration and maintenance tasks across multiple accounts and resources.

S3 uses **AES-256**



Using AWS Cognito to issue JSON Web Tokens (JWT) is incorrect because the Amazon Cognito service is primarily used for user authentication and not for providing access to your AWS resources. A **JWT is meant to be used for user authentication and session management**.

Use an active-passive failover configuration when you want a primary resource or group of resources to be available majority of the time and you want a secondary resource or group of resources to be on standby in case all the primary resources become unavailable. When responding to queries, Route 53 includes only the healthy primary resources. If all the primary resources are unhealthy, Route 53 begins to include only the healthy secondary resources in response to DNS queries.

CloudFront geo-restriction feature is primarily used to prevent users in specific geographic locations from accessing content that you're distributing through a CloudFront web distribution. It does not let you choose the resources that serve your traffic based on the geographic location of your users, unlike the Geolocation routing policy in Route 53.

By default, instances that you launch into a virtual private cloud (VPC) can't communicate with your own network. You can enable access to your network from your VPC by attaching a virtual private gateway to the VPC, creating a custom route table, updating your security group rules, and creating an AWS managed VPN connection.

An EIP to the Virtual Private Gateway is incorrect since you do not attach an EIP to a VPG. You do need a static IP for Customer Gateways.

Here are the prerequisites for routing traffic to a website that is hosted in an Amazon S3 Bucket:

- An S3 bucket that is configured to host a static website. The bucket must have the same name as your domain or subdomain. For example, if you want to use the subdomain portal.tutorialsdojo.com, the name of the bucket must be portal.tutorialsdojo.com.
- A registered domain name. You can use Route 53 as your domain registrar, or you can use a different registrar.
- Route 53 as the DNS service for the domain. If you register your domain name by using Route 53, we automatically configure Route 53 as the DNS service for the domain.

However, if you chose to use server-side encryption with customer-provided encryption keys (SSE-C), you must provide encryption key information using the following request headers:

x-amz-server-side-encryption-customer-algorithm

x-amz-server-side-encryption-customer-key

x-amz-server-side-encryption-customer-key-MD5

By default, non default subnets have the IPv4 **public addressing** attribute set to false, and default subnets have this attribute set to true. An exception is a non default subnet created by the Amazon EC2 launch instance wizard — the wizard sets the attribute to true. You can modify this attribute using the Amazon VPC console.

In this scenario, there are 5 EC2 instances that belong to the same security group that should be able to connect to the Internet. The main route table is properly configured but there is a problem connecting to one instance. Since the other four instances are working fine, we can assume that the security group and the route table are correctly configured. One possible reason for this issue is that the problematic instance does not have a public or an EIP address.

## ENI

If one of your instances serving a particular function fails, its network interface can be attached to a replacement or hot standby instance pre-configured for the same role in order to rapidly recover the service. For example, you can use a network interface as your primary or secondary network interface to a critical service such as a database instance or a NAT instance. If the instance fails, you (or more likely, the code running on your behalf) can attach the network interface to a hot standby instance.

The main requirement in the scenario is to monitor the SwapUtilization metric. Take note that **you can't use the default metrics of CloudWatch to monitor the SwapUtilization metric. To monitor custom metrics, you must install the CloudWatch agent on the EC2 instance.** After installing the CloudWatch agent, you can now collect system metrics and log files of an EC2 instance.

Here is a list of important information about EBS Volumes:

- When you create an EBS volume in an Availability Zone, it is automatically replicated within that zone to prevent data loss due to a failure of any single hardware component.

- After you create a volume, you can attach it to any EC2 instance in the same Availability Zone
- Amazon EBS Multi-Attach enables you to attach a single Provisioned IOPS SSD (io1) volume to multiple Nitro-based instances that are in the same Availability Zone. However, other EBS types are not supported.
- An EBS volume is off-instance storage that can persist independently from the life of an instance. You can specify not to terminate the EBS volume when you terminate the EC2 instance during instance creation.
- EBS volumes support live configuration changes while in production which means that you can modify the volume type, volume size, and IOPS capacity without service interruptions.
- Amazon EBS encryption uses 256-bit Advanced Encryption Standard algorithms (AES-256)
- EBS Volumes offer 99.999% SLA.

**AWS Proton** allows you to deploy any serverless or container-based application with increased efficiency, consistency, and control. You can define infrastructure standards and effective continuous delivery pipelines for your organization. Proton breaks down the infrastructure into environment and service (“infrastructure as code” templates).

As a developer, you select a standardized service template that AWS Proton uses to create a service that deploys and manages your application in a service instance. An AWS Proton service is an instantiation of a service template, which normally includes several service instances and a pipeline.

AWS Control Tower is used to simplify the **creation of new accounts with preconfigured constraints**. It isn't used to automate application deployments. Moreover, AWS Config is commonly used for monitoring the changes of AWS resources and not the custom resources for serverless or container-based applications in AWS. A combination of AWS Proton and Components is the most suitable solution for this scenario.

**Amazon SWF** (read faq) is a web service that makes it easy to coordinate work across distributed application components.

AWS Global Accelerator simply optimizes application performance by routing user traffic to the congestion-free, **redundant AWS global network instead of the public internet**.

You can use Amazon Data Lifecycle Manager (Amazon DLM) to automate the creation, retention, and deletion of snapshots taken to back up your Amazon EBS volumes. Automating snapshot management helps you to:

- Protect valuable data by enforcing a regular backup schedule.
- Retain backups as required by auditors or internal compliance.
- Reduce storage costs by deleting outdated backups.

(no thing as EBS lifecycle policy)

In Amazon RDS, failover is automatically handled so that you can resume database operations as quickly as possible without administrative intervention in the event that your primary database instance went down. **When failing over, Amazon RDS simply flips the canonical**

name record (CNAME) for your DB instance to point at the standby, which is in turn promoted to become the new primary.

**Security Groups** is for instances. **NACL** is for subnets.

Security Group	Network Access Control List
Acts as a firewall for associated Amazon EC2 instances	Acts as a firewall for associated subnets
Controls both inbound and outbound traffic at the instance level	Controls both inbound and outbound traffic at the subnet level
You can secure your VPC instances using only security groups	Network ACLs are an additional layer of defense.
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 allowed by rules)
Evaluates all rules before deciding whether to allow traffic	Evaluates rules in number order when deciding whether to allow traffic, starting with the lowest numbered rule.
Applies only to the instance that is associated to it	Applies to all instances in the subnet it is associated with
Has separate rules for inbound and outbound traffic	Has separate rules for inbound and outbound traffic
A newly created security group denies all inbound traffic by default	A newly created nACL denies all inbound traffic by default
A newly created security group has an outbound rule that allows all outbound traffic by default	A newly created nACL denies all outbound traffic default
Instances associated with a security group can't talk to each other unless you add rules allowing it	Each subnet in your VPC must be associated with a network ACL. If none is associated, the default nACL is selected.
Security groups are associated with network interfaces	You can associate a network ACL with multiple subnets; however, a subnet can be associated with only one network ACL at a time.
 <b>Tutorials Dojo</b>	

## IAM user groups

An IAM **user group** is a collection of IAM users. You can use user groups to manage permissions for those users. For example, you could have a user group called *Admins* and any user in that user group automatically has the permissions that are assigned to the user group. To assign the appropriate permissions by adding the user to that user group. When you no longer want you can remove him or her from the old user groups and add him or her to a new user group based policy. A user group is a way to attach policies to multiple users at once so they can receive the permissions from the user group. For more information about

---

## IAM roles

An IAM **role** is very similar to a user, in that it is an identity with permissions but it does not have any credentials (password or access keys) associated with it. Instead of being a user, it is a role. An IAM user can assume a role to temporarily take on different permissions. This is useful for an identity provider instead of IAM. AWS uses details passed by the identity p

Although Amazon **Kinesis Data Firehose captures and loads data in near real-time, AWS Lambda can't be set as its destination**. You can write Lambda functions and integrate it with Kinesis Data Firehose to request additional, customized processing of the data before it is sent downstream. However, this integration is primarily used for stream processing and not the actual consumption of the data stream. You have to use a Kinesis Data Stream in this scenario. RDS Storage Auto Scaling automatically scales storage capacity in response to growing database workloads, with zero downtime. The option that says: Increase the allocated storage for the DB instance is incorrect. Although this will solve the problem of low disk space, increasing the allocated storage might cause performance degradation during the change.



**If your workload is unpredictable, you can enable storage autoscaling for an Amazon RDS DB instance. To do so, you can use the Amazon RDS console, the Amazon RDS API, or the AWS CLI.**

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

When the word durability pops out, the first service that should come to your mind is Amazon S3.

The following snippet below shows how it is done using boto3 ( AWS SDK for Python ):

```
client = boto3.client('s3')
resp = client.select_object_content(
    Bucket='tdojo-bucket', # Bucket Name.
    Key='s3-select/tutorialsdojofile.csv', # Object Key.
    ExpressionType= 'SQL',
    Expression = "select \"Sample\" from s3object s where s.\"tutorialsdojofile\" in ['A', 'B']"
)
```

Hence, the correct answer is the option that says: Perform an S3 Select operation based on the bucket's name and object's key.

Amazon S3 is composed of buckets, object keys, object metadata, object tags, and many other components as shown below:

An Amazon S3 bucket name is globally unique, and the namespace is shared by all AWS accounts.

An Amazon S3 object key refers to the key name, which uniquely identifies the object in the bucket.

An Amazon S3 object metadata is a name-value pair that provides information about the object.

An Amazon S3 object tag is a key-pair value used for object tagging to categorize storage.

Kinesis Data Streams supports changes to the data record retention period of your stream. A Kinesis data stream is an ordered sequence of data records meant to be written to and read from in real-time. Data records are therefore stored in shards in your stream temporarily. The time period from when a record is added to when it is no longer accessible is called the retention period. **A Kinesis data stream stores records from 24 hours by default to a maximum of 8760 hours (365 days).**



Multi-AZ deployments	Multi-Region deployments	Read replicas
Main purpose is high availability	Main purpose is disaster recovery and local performance	Main purpose is scalability
Non-Aurora: synchronous replication; Aurora: asynchronous replication	Asynchronous replication	Asynchronous replication
Non-Aurora: only the primary instance is active; Aurora: all instances are active	All regions are accessible and can be used for reads	All read replicas are accessible and can be used for readscaling
Non-Aurora: automated backups are taken from standby; Aurora: automated backups are taken from shared storage layer	Automated backups can be taken in each region	No backups configured by default
Always span at least two Availability Zones within a single region	Each region can have a Multi-AZ deployment	Can be within an Availability Zone, Cross-AZ, or Cross-Region
Non-Aurora: database engine version upgrades happen on primary; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent in each region; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent from source instance; Aurora: all instances are updated together
Automatic failover to standby (non-Aurora) or read replica (Aurora) when a problem is detected	Aurora allows promotion of a secondary region to be the master	Can be manually promoted to a standalone database instance (non-Aurora) or to be the primary instance (Aurora)

Access logging is an optional feature of Elastic Load Balancing that is disabled by default. After you enable access logging for your load balancer, Elastic Load Balancing captures the logs and stores them in the Amazon S3 bucket that you specify as compressed files. You can disable access logging at any time.

The allowed block size in VPC is between a /16 netmask (65,536 IP addresses) and /27 netmask (32 IP addresses) is incorrect because the allowed block size in VPC is between a /16 netmask (65,536 IP addresses) and /28 netmask (16 IP addresses) and not /27 netmask. The option that says: Each subnet spans to 2 Availability Zones is incorrect because each subnet must reside entirely within one Availability Zone and cannot span zones.

You still **cannot integrate DynamoDB table with CloudFront** as these two are incompatible. The option that says: Since Auto Scaling is enabled by default, the provisioned read and write capacity will adjust automatically. Also enable DynamoDB Accelerator (DAX) to improve the performance from milliseconds to microseconds is incorrect because, by default, Auto Scaling is not enabled in a DynamoDB table which is created using the AWS CLI.

Correct answers:

- Enable DynamoDB Accelerator (DAX) and ensure that the Auto Scaling is enabled and increase the maximum provisioned read and write capacity.
- **Use API Gateway in conjunction with Lambda and turn on the caching on frequently accessed data and enable DynamoDB global replication.**

Only S3 has server side encryption, not EBS.

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.

You can use two types of VPC endpoints to access Amazon S3: gateway endpoints and interface endpoints. **A gateway endpoint is a gateway that you specify in your route table to access Amazon S3 from your VPC over the AWS network. Interface endpoints extend the functionality of gateway endpoints by using private IP addresses to route requests to Amazon S3 from within your VPC, on-premises, or from a different AWS Region.** Interface endpoints are compatible with gateway endpoints. If you have an existing gateway endpoint in the VPC, you can use both types of endpoints in the same VPC.

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 be 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.**

**AWS Backup** is a centralized backup service that makes it easy and cost-effective for you to backup your application data across AWS services in the AWS Cloud, helping you meet your business and regulatory backup compliance requirements. AWS Backup makes protecting your AWS storage volumes, databases, and file systems simple by providing a central place where you can configure and audit the AWS resources you want to backup, automate backup scheduling, set retention policies, and monitor all recent backup and restore activity. Aurora just has a backup retention period of **35 days**.

Configuring an Active-Passive Failover with Weighted Records and configuring an Active-Passive Failover with Multiple Primary and Secondary Resources are incorrect **because an Active-Passive Failover is mainly used when you want a primary resource or group of resources to be available most of the time and you want a secondary resource or group of resources to be on standby** in case all the primary resources become unavailable. In this scenario, all of your resources should be available all the time as much as possible which is why you have to use an Active-Active Failover instead.

Configuring an Active-Active Failover with One Primary and One Secondary Resource is incorrect because 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.**

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.

You are limited to running On-Demand Instances per your vCPU-based On-Demand Instance limit, purchasing **20 Reserved Instances**, and requesting Spot Instances per your dynamic Spot limit per region. New AWS accounts may start with limits that are lower than the limits described here.

Terminate the Reserved instances as soon as possible to **avoid getting billed at the on-demand price** when it expires (terminated instances can still be sold at marketplace).

Replacing the Auto Scaling group with a cluster placement group to achieve a low-latency network performance necessary for tightly-coupled node-to-node communication is incorrect because although it is true that a cluster placement group allows you to achieve a low-latency network performance, you still need to use Auto Scaling for your architecture to add more EC2 instances.

To protect your system from DDoS attack, you can do the following:

- **Use an Amazon CloudFront service for distributing both static and dynamic content.**
- Use an Application Load Balancer with Auto Scaling groups for your EC2 instances. Prevent direct Internet traffic to your Amazon RDS database by deploying it to a new private subnet.
- Set up alerts in Amazon CloudWatch to look for high Network In and CPU utilization metrics.

Adding multiple Elastic Fabric Adapters (EFA) to each EC2 instance to increase the network bandwidth is also not a viable option as this is mainly done for performance improvement, and not for DDoS attack mitigation. Moreover, you can attach only one EFA per EC2 instance. An Elastic Fabric Adapter (EFA) is a network device that you can attach to your Amazon EC2 instance to accelerate High-Performance Computing (HPC) and machine learning applications.

The option that says: Create a new AWS VPN CloudHub. Set up a Virtual Private Network (VPN) connection for additional AWS accounts is incorrect because a VPN connection is not capable of providing consistent and dedicated access to the on-premises network services.

**Take note that a VPN connection traverses the public Internet and doesn't use a dedicated connection.** (like a peer to peer/transit does)

Another main distinction is that cloudhub is a hub and spoke model while transit/VPN is as aforementioned a dedicated line.

Remember that if you configure CloudFront to serve HTTPS requests using dedicated IP addresses, you incur an additional monthly charge. The charge begins when you associate your SSL/TLS certificate with your CloudFront distribution. You can just simply upload the certificates to the ALB and use SNI to handle multiple domains in a cost-effective manner.

Glacier -> use expedited and provisioned for fastest retrieval. Glacier Select is just for filtering.

Keep in mind that an EC2 instance has an underlying physical host computer. **If the instance is stopped, AWS usually moves the instance to a new host computer.** Your instance may stay on the same host computer if there are no problems with the host computer. In addition, its Elastic IP address is disassociated from the instance if it is an EC2-Classic instance. Otherwise, if it is an EC2-VPC instance, the Elastic IP address remains associated. The option that says: The ENI (Elastic Network Interface) is detached is incorrect because the ENI will stay attached even if you stopped your EC2 instance.

1024–65535 are ephemeral/temporary port numbers just for the duration of a transport layer communication session.

**Origin Shield**, a centralized caching layer that helps increase your cache hit ratio to reduce the load on your origin. Origin Shield also decreases your origin operating costs by collapsing requests across regions so as few as one request goes to your origin per object.

AWS Virtual Private Network (**VPN**) is incorrect because this service is **mainly used for establishing encryption connections from an on-premises network to AWS.**

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.

**Storage Gateway can't interact with S3.**

The option that says: Enable Amazon S3 Transfer Acceleration on the target S3 bucket is incorrect because this S3 feature is not suitable for large-scale data migration. Enabling this feature won't always guarantee faster data transfer as it's only beneficial for long-distance transfer to and from your Amazon S3 buckets.

The option that says: **Integrate AWS Storage Gateway File Gateway** with the on-premises data center is incorrect because this service is mostly used for building hybrid cloud solutions **where you still need on-premises access to unlimited cloud storage.** Based on the scenario, this service is not the best option because you would still rely on the existing low bandwidth internet connection.

Volume Gateway – provides **cloud-backed storage volumes that you can mount as iSCSI devices from your on-premises application servers.**

**Target groups are primarily used in ELBs and not in Auto Scaling.** The scenario didn't mention that the architecture has a load balancer. Therefore, you should be updating your launch configuration, not the target group.

#### Tutorial Dojo Exam 4

Amazon SWF provides useful guarantees around task assignments. **It ensures that a task is never duplicated and is assigned only once.** Thus, even though you may have multiple workers for a particular activity type (or a number of instances of a decider), Amazon SWF will give a specific task to only one worker (or one decider instance). Additionally, Amazon SWF keeps at most one decision task outstanding at a time for a workflow execution. Thus, you can run multiple decider instances without worrying about two instances operating on the same execution simultaneously. These facilities enable you to coordinate your workflow without worrying about duplicate, lost, or conflicting tasks.

The main issue in this scenario is that the order management system produces duplicate orders at times. Since the company is using SQS, there is a possibility that a message can have a duplicate in case an EC2 instance failed to delete the already processed message. To prevent this issue from happening, you have to use Amazon Simple Workflow service instead of SQS.

Config can check for **password compliance** as well.

#### Unified Cloudwatch - centralized logs (Cloudwatch Log Insights)

One of the differences between Fault Tolerance and High Availability is that the former refers to the minimum number of running instances. For example, you have a system that requires a minimum of 4 running instances and currently has 6 running instances deployed in two Availability Zones. There was a component failure in one of the Availability Zones which knocks out 3 instances. In this case, the system can still be regarded as Highly Available since there are still instances running that can accommodate the requests. However, it is not Fault-Tolerant since the required minimum of four instances has not been met.

So you need at **minimum** at least the required number of instances running all the time (even if redundant).

**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.

CloudTrail is incorrect because this is primarily used for IT audits and API logging of all of your AWS resources. It does not have the capability to trace and analyze user requests as they travel through your Amazon API Gateway APIs, unlike **AWS X-Ray**.

S3 one zone/normal IA **both** have millisecond retrieval times.

CloudFormation enables modeling, provisioning, **and version-controlling of your entire AWS infrastructure** (cloudformation itself is **free**)

Redshift Spectrum is incorrect. Although Amazon Redshift Spectrum provides a similar in-query functionality like S3 Select, **this service is more suitable for querying your data from the Redshift external tables hosted in S3**. The Redshift queries are run on your cluster resources against local disk. Redshift Spectrum queries run using per-query scale-out resources against data in S3 which can entail additional costs compared with S3 Select.

With **Amazon S3 Select**, you can use simple structured query language (SQL) statements to filter the contents of Amazon S3 objects and retrieve just the subset of data that you need. By using Amazon S3 Select to filter this data, you can reduce the amount of data that Amazon S3 transfers, which reduces the cost and latency to retrieve this data.

Feature	APPLICATION LOAD BALANCER	NETWORK LOAD BALANCER	CLASSIC LOAD BALANCER
Protocols	HTTP, HTTPS	TCP, UDP, TLS	TCP, SSL/TLS, HTTP, HTTPS
Platforms	VPC	VPC	EC2-Classic, VPC
Health checks	✓	✓	✓
CloudWatch metrics	✓	✓	✓
Logging	✓	✓	✓
Zonal fail-over	✓	✓	✓
Connection draining (deregistration delay)	✓		✓
Load Balancing to multiple ports on the same instance	✓	✓	
IP addresses as targets	✓	✓ (TCP, TLS)	
Load balancer deletion protection	✓	✓	
Configurable idle connection timeout	✓		✓
Cross-zone load balancing	✓	✓	✓
Sticky sessions	✓	✓	✓
Static IP		✓	
Elastic IP address		✓	
Preserve Source IP address		✓	
Resource-based IAM Permissions	✓	✓	✓
Tag-based IAM permissions	✓	✓	
Slow start	✓		
WebSockets	✓	✓	
PrivateLink Support		✓ (TCP, TLS)	
Source IP address CIDR-based routing	✓		

Tutorials Dojo

You can only create a private virtual interface to a Direct Connect gateway and not a public virtual interface. Using a link aggregation group (LAG) is also irrelevant in this scenario because

it is just a logical interface that uses the Link Aggregation Control Protocol (LACP) to aggregate multiple connections at a single AWS Direct Connect endpoint, allowing you to treat them as a single, managed connection.

Amazon S3 is integrated with AWS CloudTrail, a service that provides a record of actions taken by a user, role, or an AWS service in Amazon S3. CloudTrail captures a subset of API calls for Amazon S3 as events, including calls from the Amazon S3 console and code calls to the Amazon S3 APIs.

AWS CloudTrail logs provide a record of actions taken by a user, role, or an AWS service in Amazon S3, while **Amazon S3 server access logs provide detailed records for the requests that are made to an S3 bucket.**

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. 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.

AWS WAF is incorrect because this is a web application firewall service that helps protect your web apps from common exploits that could affect app availability, compromise security, or consume excessive resources. Although this can help you against DDoS attacks, AWS WAF alone is not enough to fully protect your VPC. You still need to use AWS Shield in this scenario. AWS **Firewall Manager** is incorrect because this just **simplifies your AWS WAF administration and maintenance tasks across multiple accounts and resources.**

Install the CloudWatch unified agent to the EC2 instances. Set up a custom parameter in AWS Systems Manager Parameter Store with the CloudWatch agent configuration to create an aggregated metric on memory usage percentage. Scale the Auto Scaling group based on the aggregated metric.



<b>C: Compute Optimized Instances</b>	<b>Cost-effective high performance at a low price per compute ratio</b>
<b>D: Storage Optimized Instances</b>	High disk throughput
<b>G: Accelerated Computing Instances</b>	Graphics-intensive GPU instances
<b>H: Storage Optimized Instances</b>	HDD-based local storage for high disk throughput
<b>I: Storage Optimized Instances</b>	High storage instances, low latency, high random I/O performance, high sequential read throughput, and high IOPS
<b>M: General Purpose Instances</b>	Fixed performance
<b>P: Accelerated Computing Instances</b>	General purpose GPU instances
<b>F: Accelerated Computing Instances</b>	Reconfigurable FPGA instances
<b>R: Memory Optimized Instances</b>	Memory-intensive applications
<b>T: General Purpose Instances</b>	Burstable performance instances
<b>X: Memory Optimized Instances</b>	Large-scale, enterprise-class, in-memory applications, and high-performance databases

Lambda@Edge is a feature of Amazon CloudFront that **lets you run code closer to users of your application**, which improves performance and reduces latency. With Lambda@Edge, you don't have to provision or manage infrastructure in multiple locations around the world. You pay only for the compute time you consume – there is no charge when your code is not running.

You **pay for all bandwidth into and out of Amazon S3**, except for the following:

- Data transferred in from the Internet.
- **Data transferred out to an Amazon EC2 instance, when the instance is in the same AWS Region** as the S3 bucket (including to a different account in the same AWS region).
- Data transferred out to Amazon CloudFront.

Snowball is for anything on-premises. Even snowball edge.

If you got your certificate from a third-party CA, import the certificate into ACM or upload it to the IAM certificate store. Hence, AWS Certificate Manager and IAM certificate store are the correct answers.

Both S3 and EBS gives the availability of 99.99%, but the only difference that occurs is that S3 is accessed via the internet using API's and EBS is accessed by the single instance attached to EBS.

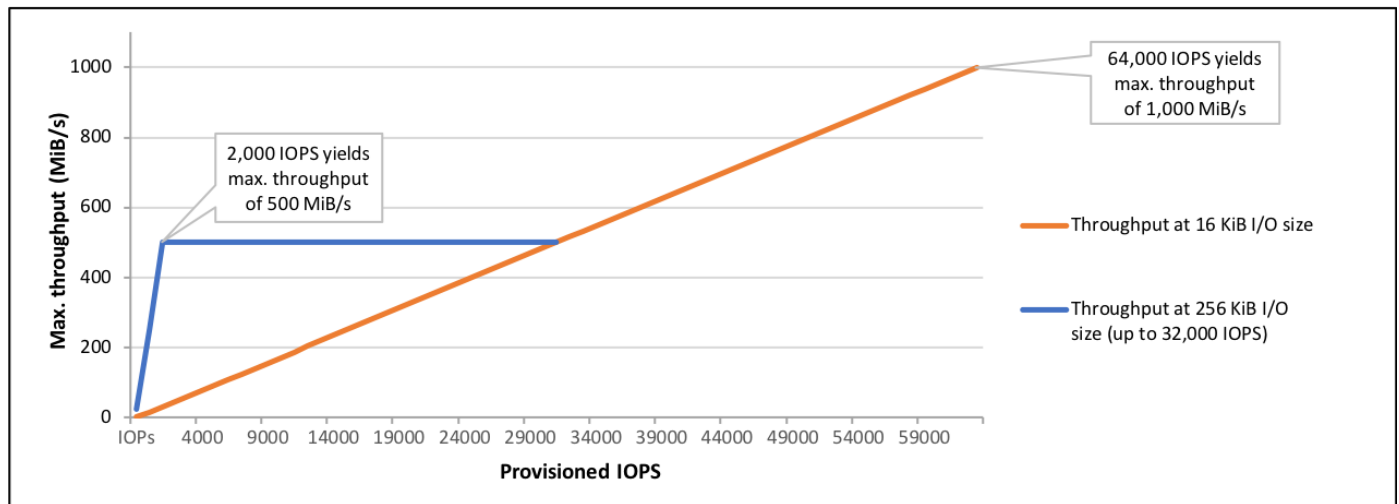
Storage Need	Solution	AWS Services
<b>Temporary storage</b>	Consider using local instance store volumes for needs such as scratch disks, buffers, queues, and caches.	<a href="#">Amazon Local Instance Store</a>
<b>Multi-instance storage</b>	Amazon EBS volumes can only be attached to one EC2 instance at a time. If you need multiple EC2 instances accessing volume data at the same time, consider using Amazon EFS as a file system.	<a href="#">Amazon EFS</a>
<b>Highly durable storage</b>	If you need very highly durable storage, use S3 or Amazon EFS. Amazon S3 Standard storage is designed for 99.999999999 percent (11 nines) annual durability per object. You can even decide to take a snapshot of the EBS volumes. Such a snapshot then gets saved in Amazon S3, thus providing you the durability of Amazon S3. For more information on EBS durability, see the <a href="#">Durability and Availability</a> section. EFS is designed for high durability and high availability, with data stored in multiple Availability Zones within an AWS Region.	<a href="#">Amazon S3</a> <a href="#">Amazon EFS</a>
<b>Static data or web content</b>	If your data doesn't change that often, Amazon S3 might represent a more cost-effective and scalable solution for storing this fixed information. Also, web content served out of Amazon EBS requires a web server running on Amazon EC2; in contrast, you can deliver web content directly out of Amazon S3 or from multiple EC2 instances using Amazon EFS.	<a href="#">Amazon S3</a> <a href="#">Amazon EFS</a>

The option that says: Configure RAID 1 in multiple instance volumes is incorrect because RAID 1 configuration is used for data mirroring. You need to configure RAID 0 to improve the performance of your storage volumes.

Provisioned IOPS SSD (io1) volumes are designed to meet the needs of I/O-intensive workloads, particularly database workloads, that are sensitive to storage performance and consistency. Unlike gp2, which uses a bucket and credit model to calculate performance, an io1 volume allows you to specify a consistent IOPS rate when you create the volume, and Amazon EBS delivers within 10 percent of the provisioned IOPS performance 99.9 percent of the time over a given year.

An io1 volume can range in size from 4 GiB to 16 TiB. You can provision from 100 IOPS up to 64,000 IOPS per volume on Nitro system instance families and up to 32,000 on other instance families. The maximum ratio of provisioned IOPS to requested volume size (in GiB) is 50:1.

For example, a 100 GiB volume can be provisioned with up to 5,000 IOPS. On a supported instance type, any volume 1,280 GiB in size or greater allows provisioning up to the 64,000 IOPS maximum ( $50 \times 1,280 \text{ GiB} = 64,000$ ).



An io1 volume provisioned with up to 32,000 IOPS supports a maximum I/O size of 256 KiB and yields as much as 500 MiB/s of throughput. With the I/O size at the maximum, peak throughput is reached at 2,000 IOPS. A volume provisioned with more than 32,000 IOPS (up to the cap of 64,000 IOPS) supports a maximum I/O size of 16 KiB and yields as much as 1,000 MiB/s of throughput.

The volume queue length is the number of pending I/O requests for a device. Latency is the true end-to-end client time of an I/O operation, in other words, the time elapsed between sending an I/O to EBS and receiving an acknowledgment from EBS that the I/O read or write is complete. Queue length must be correctly calibrated with I/O size and latency to avoid creating bottlenecks either on the guest operating system or on the network link to EBS.

Optimal queue length varies for each workload, depending on your particular application's sensitivity to IOPS and latency. If your workload is not delivering enough I/O requests to fully use the performance available to your EBS volume then your volume might not deliver the IOPS or throughput that you have provisioned.

Transaction-intensive applications are sensitive to increased I/O latency and are well-suited for SSD-backed io1 and gp2 volumes. You can maintain high IOPS while keeping latency down by maintaining a low queue length and a high number of IOPS available to the volume. Consistently driving more IOPS to a volume than it has available can cause increased I/O latency.

Throughput-intensive applications are less sensitive to increased I/O latency, and are well-suited for HDD-backed st1 and sc1 volumes. You can maintain high throughput to HDD-backed volumes by maintaining a high queue length when performing large, sequential I/O.

Therefore, for instance, a 10 GiB volume can be provisioned with up to 500 IOPS. Any volume 640 GiB in size or greater allows provisioning up to a maximum of 32,000 IOPS ( $50 \times 640 \text{ GiB} = 32,000$ ). Hence, the correct answer is to set the IOPS to 500 then maintain a low queue length.

Setting the IOPS to 400 then maintaining a low queue length is incorrect because although a value of 400 is an acceptable value, it is not the maximum value for the IOPS. You will not fully utilize the available IOPS that the volume can offer if you just set it to 400.

The options that say: Set the IOPS to 600 then maintain a high queue length and Set the IOPS to 800 then maintain a low queue length are both incorrect because the maximum IOPS for the 10 GiB volume is only 500. Therefore, any value greater than the maximum amount, such as 600 or 800, is wrong. Moreover, you should keep the latency down by maintaining a low queue length, and not higher.

#### References:

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-io-characteristics.html>

—

A company has a two-tier environment in its on-premises data center which is composed of an application tier and database tier. You are instructed to migrate their environment to the AWS cloud, and to design the subnets in their VPC with the following requirements:

1. There is an application load balancer that would distribute the incoming traffic among the servers in the application tier.
2. The application tier and the database tier must not be accessible from the public Internet. The application tier should only accept traffic coming from the load balancer.
3. The database tier contains very sensitive data. It must not share the same subnet with other AWS resources and its custom route table with other instances in the environment.
4. The environment must be highly available and scalable to handle a surge of incoming traffic over the Internet.

How many subnets should you create to meet the above requirements?

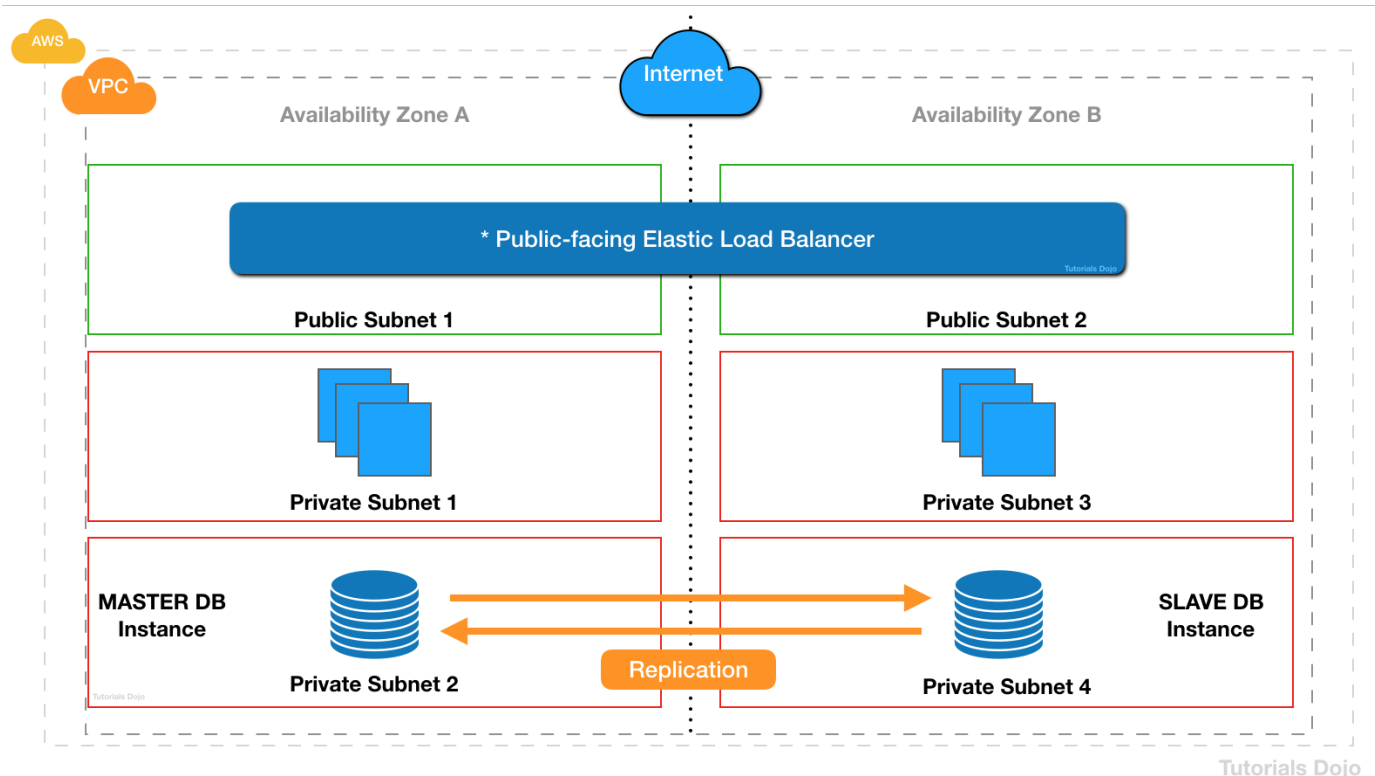
The given scenario indicated 4 requirements that should be met in order to successfully migrate their two-tier environment from their on-premises data center to AWS Cloud. The first

requirement means that you have to use an application load balancer (ALB) to distribute the incoming traffic to your application servers.

The second requirement specifies that both your application and database tier should not be accessible from the public Internet. This means that you could create a single private subnet for both of your application and database tier. However, the third requirement mentioned that the database tier should not share the same subnet with other AWS resources to protect its sensitive data. This means that you should provision one private subnet for your application tier and another private subnet for your database tier.

The last requirement alludes to the need for using at least two Availability Zones to achieve high availability. This means that you have to distribute your application servers to two AZs as well as your database which can be set up with a master-slave configuration to properly replicate the data between two zones.

If you have more than one private subnet in the same Availability Zone that contains instances that need to be registered with the load balancer, you only need to create one public subnet. You need only one public subnet per Availability Zone; you can add the private instances in all the private subnets that reside in that particular Availability Zone.



Since you have a public internet-facing load balancer that has a group of backend Amazon EC2 instances that are deployed in a private subnet, you must create the corresponding public subnets in the same Availability Zones. This new public subnet is on top of the private subnet

that is used by your private EC2 instances. Lastly, you should associate these public subnets to the Internet-facing load balancer to complete the setup.

To summarize, we need to have one private subnet for the application tier and another one for the database tier. We then need to create another public subnet in the same Availability Zone where the private EC2 instances are hosted, in order to properly connect the public Internet-facing load balancer to your instances. This means that we have to use a total of 3 subnets consisting of 2 private subnets and 1 public subnet.

To meet the requirement of high availability, we have to deploy the stack to two Availability Zones. This means that you have to double the number of subnets you are using. Take note as well that you must create the corresponding public subnet in the same Availability Zone of your private EC2 servers in order for it to properly communicate with the load balancer.

Hence, the correct answer is 6 subnets.

—

### **Tutorial Dojo Exam 5**

While it is completing, an in-progress snapshot is not affected by ongoing reads and writes to the volume hence, you can still use the EBS volume normally.

In this scenario, the technology company is looking for a storage solution that can store data subsets and not the entire data set (as it was mentioned that they only need to store up to 10% of the data). The correct answer is to set up a Cached Volume Gateway in AWS Storage Gateway.

By using cached volumes, you can use Amazon S3 as your primary storage. The gateway acts as a local cache. Cached volumes minimize the need to scale your on-premises storage. You can create storage volumes and attach them to your on-premises application servers. When you write to the volumes, the data is stored in the cache. When you read data, it is retrieved from the cache and then from Amazon S3. Recently read data is stored in your on-premises storage gateway's cache and is available for future reads.

Cached volumes can range from 1 GiB to 32 TiB in size and must be attached to a single on-premises application server. A single storage gateway can support up to 32 volumes for a total maximum storage volume of 1024 TiB.

In the cached volumes solution, AWS Storage Gateway stores all your data in Amazon S3. The correct answer is: **Volume Gateway in cached mode.**

File Gateway is incorrect because the scenario requires you to mount volumes as iSCSI devices. File Gateway is used to store and retrieve Amazon S3 objects through NFS and SMB protocols.

You won't get billed if you use a Gateway VPC endpoint for your Amazon S3 bucket, unlike an Interface VPC endpoint that is billed for hourly usage and data processing charges.

Gateway endpoints for Amazon	
	In bot
Use Amazon S3 public IP address	
Does not allow access from on prem	
Does not allow access from another AW	
Not billed	

An interface endpoint is an elastic network interface with a private IP address that serves as an entry point for traffic destined to a supported service. A gateway endpoint is a gateway that is a target for a specified route in your route table, used for traffic destined to a supported AWS service.

- Use Gateway Endpoint if the AWS service is either DynamoDB or S3.
- Use Interface Endpoint for everything else.

Network Load balancer doesn't have Weighted Target Groups to divert the traffic between the on-premises and AWS-hosted application.

When you create a target group in your Application Load Balancer, you specify its target type. This determines the type of target you specify when registering with this target group. You can select the following target types:

1. instance – The targets are specified by instance ID.
2. ip – The targets are IP addresses.
3. Lambda – The target is a Lambda function.

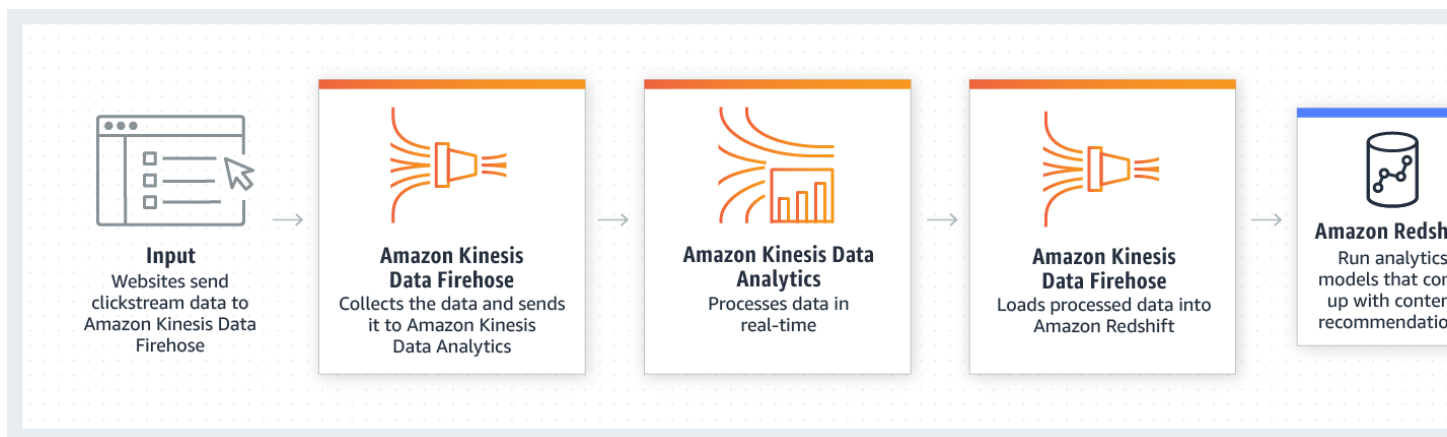


Since the scenario asks you to create a short-lived authentication token to access an Amazon RDS database, you can use an IAM database authentication when connecting to a database instance. Authentication is handled by **AWSAuthenticationPlugin**—an AWS-provided plugin that works seamlessly with IAM to authenticate your IAM users.

Management Events provide visibility into management operations that are performed on resources in your AWS account. These are also known as control plane operations.

Management events can also include non-API events that occur in your account.

Data Events, on the other hand, provide visibility into the resource operations performed on or within a resource. These are also known as data plane operations. It allows granular control of data event logging with advanced event selectors. You can currently log data events on different resource types such as Amazon S3 object-level API activity (e.g. GetObject, DeleteObject, and PutObject API operations), AWS Lambda function execution activity (the Invoke API), DynamoDB Item actions, and many more.



Using a password stored in CloudHSM is incorrect as you only store keys in CloudHSM and not passwords.

Short polling -> sampling



Long polling -> pub sub type deal, client will inform when a new msg comes and you can set

ReceiveMessageWaitTimeSeconds

Quick facts about SQS Long Polling:

- Long polling helps reduce your cost of using Amazon SQS by reducing the number of empty responses when there are no messages available to return in reply to a ReceiveMessage request sent to an Amazon SQS queue and eliminating false empty responses when messages are available in the queue but aren't included in the response.
- Long polling reduces the number of empty responses by allowing Amazon SQS to wait until a message is available in the queue before sending a response. Unless the connection times out, the response to the ReceiveMessage request contains at least one of the available messages, up to the maximum number of messages specified in the ReceiveMessage action.

– Long polling eliminates false empty responses by querying all (rather than a limited number) of the servers. Long polling returns messages as soon any message becomes available.

FEATURES	SSD Solid State Drive	HDD Hard Disk Drive
Best for workloads with:	<i>small, random</i> I/O operations	<i>large, sequential</i> I/O operations
Can be used as a bootable volume?	Yes	No
Suitable Use Cases	<ul style="list-style-type: none"><li>- Best for <b>transactional workloads</b></li><li>- Critical business applications that require sustained IOPS performance</li><li>- Large database workloads such as MongoDB, Oracle, Microsoft SQL Server and many others...</li></ul>	<ul style="list-style-type: none"><li>- Best for <b>large streaming workloads</b> requiring consistent, fast throughput at a low price</li><li>- Big data, Data warehouses, Log processing</li><li>- Throughput-oriented storage for large volumes of data that is <b>infrequently</b> accessed</li></ul>
Cost	moderate / high 	low 
Dominant Performance Attribute	IOPS	Throughput (MiB/s)



Only SSDs bootable

The aws-auth ConfigMap is automatically created and applied to your cluster when you create a managed node group or when you create a node group using eksctl. It is initially created to allow nodes to join your cluster, but you also use this ConfigMap to add role-based access control (RBAC) access to IAM users and roles.

Since there is a time constraint in transitioning objects in S3, you can only change the storage class of your objects from S3 Standard storage class to STANDARD\_IA or ONEZONE\_IA storage after 30 days. This limitation does not apply on INTELLIGENT\_TIERING, GLACIER, and DEEP\_ARCHIVE storage class.

In addition, the requirement says that the media assets should be fetched in a matter of minutes for a surprise annual data audit. This means that the retrieval will only happen once a year. You can use expedited retrievals in Glacier which will allow you to quickly access your data (within 1–5 minutes) when occasional urgent requests for a subset of archives are required.

**In this scenario, you can set a lifecycle policy in the bucket to transition to S3 – Standard IA after 30 days or alternatively, you can directly transition your data to Glacier after one week (7 days).**

Amazon **SQS automatically deletes messages** that have been in a queue for more than the maximum message retention period. The **default message retention period is 4 days**. Since the queue is configured to the default settings and the batch job application only processes the messages once a week, the messages that are in the queue for more than 4 days are deleted. This is the root cause of the issue.

Amazon Redshift is incorrect because this is primarily used for OLAP applications and not for OLTP. Moreover, it doesn't scale automatically to handle the exponential growth of the database.

**Redshift OLAP ANALYTICAL**

**Aurora OLTP TRANSACTIONAL (A to T not A to A)**

**Aurora scales up to 64TB.**

The option that says: In the Application Load Balancer, create a listener rule that explicitly allows requests from approved IP addresses is incorrect because a listener rule just checks for connection requests using the protocol and port that you configure. It only determines how the load balancer routes the requests to its registered targets.

If you have an Amazon Aurora Replica in the same or a different Availability Zone, when failing over, Amazon Aurora flips the canonical name record (CNAME) for your DB Instance to point at the healthy replica, which in turn is promoted to become the new primary. Start-to-finish, failover typically completes within 30 seconds. **Aurora points failover to same AZ.** .

If you are running Aurora Serverless and the DB instance or AZ become unavailable, Aurora will automatically recreate the DB instance in a different AZ. **Serverless Aurora to different AZ.**

(Also Aurora flips the CNAME not the A record).

S3 scales performance automatically.

AWS workspaces go with Directory Services.

A: An Amazon WorkSpace is a cloud-based virtual desktop that can act as a replacement for a traditional desktop. A WorkSpace is available as a bundle of operating system, compute resources, storage space, and software applications that allow a user to perform day-to-day tasks just like using a traditional desktop.

Network Access Analyzer is a feature of VPC that reports on unintended access to your AWS resources based on the security and compliance that you set.

The option that says: Provide permissions to the users via the AWS Resource Access Manager (RAM) service to only access EC2 instances that are used for production or development is incorrect because the **AWS Resource Access Manager (RAM) is primarily used to securely share your resources across AWS accounts or within your Organization and not on a single AWS account.** You also have to set up a custom IAM Policy in order for this to work.

AWS Trusted Advisor is an online tool that provides you real-time guidance to help you provision your resources following AWS best practices. It inspects your AWS environment and

makes recommendations for saving money, improving system performance and reliability, or closing security gaps.

Trusted Advisor includes an ever-expanding list of checks in the following five categories:

**Cost Optimization** – recommendations that can potentially save you money by highlighting unused resources and opportunities to reduce your bill.

**Security** – identification of security settings that could make your AWS solution less secure.

**Fault Tolerance** – recommendations that help increase the resiliency of your AWS solution by highlighting redundancy shortfalls, current service limits, and over-utilized resources.

**Performance** – recommendations that can help to improve the speed and responsiveness of your applications.

**Service Limits** – recommendations that will tell you when service usage is more than 80% of the service limit.

Amazon Inspector is incorrect because it is just an automated security assessment service that helps improve the security and compliance of applications deployed on AWS. Amazon Inspector automatically assesses applications for exposure, vulnerabilities, and deviations from best practices.

Target Scaling is for one target, while step scaling can be for a whole group of metrics.

**EFA is not supported by Windows, it becomes a regular ENA.** (Fuck becomes Nut if not windows)

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.

## **Tutorial Dojo Exam 6**

### **Second run**

Volume Gateway stored volume saves everything

RAM (resource access manager) is used to share resources across accounts, not on a single account

EC2 by itself-> cpu disk network | Cloudwatch ->

- Memory utilization
- Disk swap utilization
- Disk space utilization
- Page file utilization
- Log collection

RPO -> data lost

RTO -> time lost

RAID 0 is striping. RAID 1 is mirroring

Redis > Memcache when you need replication/snapshots/pubsub etc

Name Server (NS) -> DNS queries

WAF can block region wise

RDS events only provide operational events such as DB instance events, DB parameter group events, DB security group events, and DB snapshot events. What we need in the scenario is to capture data-modifying events (INSERT, DELETE, UPDATE) which can be achieved thru native functions or stored procedures.

AWS direct connect is only for on-premises to cloud. Cloudhub is connections between two remote places. Use WAF to block dynamically changing IPs hijacking your app

Lookup S3 lifecycle dates (30 days for IA? 3 days for direct glacier)

SQS -> 1 minute to 14 days. Default 4 days

Kinesis Data Streams holds by default to 1 day but can hold up to 365 days

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.

Route53 geographic routing -> serve traffic based on geographic location

Cloudfront geo-restriction -> restricts based on location

Cloudwatch -> 5 min reports (1 min with detailed reporting)

Data Lifecycle manager manages snapshots

Security Groups are stateful (no need to configured outgoing as well) while Network ACL are stateless (configured incoming/outgoing seperately)

ALB does gRPC

sns and cloudwatch to send email

in sqs requests have to be manually deleted

AWS Direct Connect is a dedicated network connection from your datacenter to AWS. Due to its high cost, you should only invest in Direct Connect if you require continuous replication and connectivity between AWS and your datacenter. If you're making a one-time move to AWS, building a Direct Connect is a waste.

Alias -> within VPCs

Non-Alias -> outside of VPC

type AAAA -> ipv6

type A -> ipv4

AWS Firewall Manager just \*manages\* WAF and Shield across multiple resources

If >10TB and takes >week then use Snowball

AWS direct connect doesn't make sense for just one time transfers

Fanout starts with SNS and goes to multiple SQS queues

You can't create a VPC peering for your on-premises network and AWS VPC

sc1 infrequently accessed. st1 frequently accessed.

relationship databases -> transactional/analytics

noSQL -> transactional

aurora -> transactional

redshift -> analytical

OLAP is when you need to aggregate historical data for analysis. Transactional is processing real time data.

kinesis 24 hours, default backup doesn't exceed 35 days without aws backup

sqs 1 minute to 14 days. Default 4 days. You don't have to delete sqs messages like sns.

sns stay in queue for 3 days 72 hrs

ebs non-root persists? To preserve the root volume when an instance terminates, change the DeleteOnTermination attribute for the root volume to False.

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

mysql rds allows for autoscaling as well (dont 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.

Amazon CloudWatch Application Insights facilitates observability for your applications and underlying AWS resources. It helps you set up the best monitors for your application resources to continuously analyze data for signs of problems with your applications. Application Insights, which is powered by SageMaker and other AWS technologies, provides automated dashboards that show potential problems with monitored applications, which help you to quickly isolate ongoing issues with your applications and infrastructure. The enhanced visibility into the health of your applications that Application Insights provides helps reduce the “mean time to repair” (MTTR) to troubleshoot your application issues.

One Subnet -> One AZ

Subnet is automatically associated with main route table for VPC

Without AWS backup max retention period is 35 days

Your VPC has an implicit router and you use route tables to control where network traffic is directed. Each subnet in your VPC must be associated with a route table, which controls the routing for the subnet (subnet route table). You can explicitly associate a subnet with a particular route table. Otherwise, the subnet is implicitly associated with the main route table.

Use file gateway to access files, not volume gateway (eg S3 APIs to access files)

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.)

You can modify RDS DB settings for provisioned IOPS and storage autoscaling

Firehose can send to databases but not to lambda, also supports SNS now

Storage Gateway has local caching

Lambda execution roles are used to give permission to other resources

Cloudwatch insights detects unusual API activities, log file validation tracks validation of log files and if they have been tampered with

DynamoDB Transactions -> processing financial transactions, fulfilling/managing orders, building multiplayer game engines, coordinating actions across distributed components and services

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.

User pools are for authentication (identity verification). With a user pool, your app users can sign in through the user pool or federate through a third-party identity provider (IdP). Identity pools are for authorization (access control). You can use identity pools to create unique identities for users and give them access to other AWS services.

The difference between blue-green deployments and A/B testing is A/B testing is for measuring functionality in the app. Blue-green deployments is about releasing new software safely and rolling back predictably

EventBridge delivers a stream of real-time data from event sources such as Zendesk or Shopify to targets like AWS Lambda and other SaaS applications. You can set up routing rules to determine where to send your data to build application architectures that react in real-time to your data sources with event publisher and consumer completely decoupled.

Step Functions > SWF (no need to manage a infrastructure)

Datasync just syncs over data.

Storage Gateways:

file gateway -> access files on s3

volume gateway -> snapshots, data stored locally and synced to s3

tape gateway -> send over directly to glacier

AWS Site-to-Site VPN enables you to securely connect your on-premises network or branch office site to your Amazon VPC.

AWS Client VPN enables you to securely connect users to AWS or on-premises networks

The difference between Amazon Inspector and Amazon GuardDuty is that the former "checks what happens when you actually get an attack" and the latter "analyzes the actual logs to check if a threat exists". The purpose of Amazon Inspector is to test whether you are addressing common security risks in the target AWS. 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 in AWS.

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

SCP can override IAM

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.

VPC gateway endpoint support s3 and dynamoDB (rest have interface endpoint).

MQ is when you are moving an existing system over to cloud, if starting over then use SNS/SQS

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.

Provisioned Aurora cluster -> opposite of serverless

s3/lustre for cold/hot storage

sqs inflight limit 120,000 and 20,000 for fifo

Geoproximity has bias/geographic region, geolocation doesn't

Workspaces can do virtual desktops

Pilotlight failovers: Route53, Global Accelerator (anycastIP, give an IP to a region), Cloudfront

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.

bastion host, don't allow /0 (use specific IP ranges)

EFA (HPC) not supported on Windows

Either create an encrypted volume and copy data to it or take a snapshot, encrypt it, and create a new encrypted volume from the snapshot

redis stores data persistently (along with snapshots/clustering) , memcached does not

You can maintain high IOPS while keeping latency down by maintaining a low queue length and a high number of IOPS available to the volume.  
You can maintain high throughput to HDD-backed

volumes by maintaining a high queue length when performing large, sequential I/O

throughput -> high queue length. IOPS -> low queue length.

Access pattern: S3 access through gateway endpoints is supported only for resources in a specific VPC to which the endpoint is associated. S3 gateway endpoints do not currently support access from resources in a different Region, different VPC, or from an on-premises (non-AWS) environment. However, if you're willing to manage a complex custom architecture, you can use proxies. In all those scenarios, where access is from resources external to VPC, S3 interface endpoints access S3 in a secure way.

Athena -> query data across multiple buckets

Global Accelerator -> non-HTTP such as gaming or IOT or VoIP. Handles UDP and TCP as well. Cloudfront does images/videos/static things

CloudFront uses Edge Locations to cache content while Global Accelerator uses Edge Locations to find an optimal pathway to the nearest regional endpoint

Q. Limit the maximum number of requests from a single IP address.

A. Create a rate-based rule in AWS WAF and set the rate limit.

Q. Secure the web application by allowing multiple domains to serve SSL traffic over the same IP address.

A. Use AWS Certificate Manager to generate an SSL certificate. Associate the certificate to the CloudFront distribution and enable Server Name Indication (SNI).

EMR launches in the same AZ (HPC, is hosted Hadoop cluster running on EC2 and S3)

RAID 0 = 0 striping – data is written across multiple disks and increases performance but no redundancy

RAID 1 = 1 mirroring – creates 2 copies of the data but does not increase performance, only redundancy

To configure your Auto Scaling group to scale based on a schedule, you create a scheduled action. The scheduled action tells Amazon EC2 Auto Scaling to perform a scaling action at specified times. You cannot use scheduled action to carry out custom actions when the Auto Scaling group launches or terminates an instance.

Udemy Test 3 7 11 14 16 18 19 20 24 27 28 34 38 41 42 43 50 53 58

You cannot directly copy from Snowball Edge into Glacier without a Lifecycle Policy.

Aurora doesn't have caching, ElastiCache does

Cross Region Replication only works for NEW data, not for existing data. Use S3 sync

Lambda can access other account's data (IAM roles)

S3 storage analyzer doesn't go beyond Standard IA (doesn't recommend glacier or onezone IA)

two main components of Amazon Cognito are user pools and identity pools. Identity pools provide AWS credentials to grant your users access to other AWS services. To enable users in your user pool to access AWS resources, you can configure an identity pool to exchange user pool tokens for AWS credentials. So, identity pools aren't an authentication mechanism in themselves and hence aren't a choice for this use case

ECS with EC2 launch type is charged based on EC2 instances and EBS volumes used. ECS with Fargate launch type is charged based on vCPU and memory resources that the containerized application requests

AWS Glue - AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics. AWS Glue job is meant to be used for batch ETL data processing. AWS Glue does not offer the same storage and processing speed as FSx for Lustre. So it is not the right fit for the given high-performance workflow scenario

Multipart upload provides improved throughput, therefore it facilitates faster file uploads.  
Direct Connect takes several months

Create a new IAM role with the required permissions to access the resources in the production environment. The users can then assume this IAM role while accessing the resources from the production environment

Don't need Elastic IP for ALB

Can't share VPC itself, you share subnets with VPC sharing

SCHEDULED ACTIONS ONLY DO SCALING

Cloudwatch alarms can start/stop instances

GuardDuty -> CloudTrail, VPC Flow Logs, DNS Logs (threat detection system)

AWS Detective -> puts logs all in one place

Trusted Advisor -> sees if services follow core AWS values

Macie -> protects sensitive data

Inspector -> compliance, network connectivity/risk

launch template -> ASGs, launch configuration -> EC2s

You can stream data from Amazon S3 to Amazon Kinesis Data Streams using AWS DMS

S3 can directly invoke event notification

SQS/SNS now has VPC endpoint

AWS Inspector leverages agents installed on EC2 instances and performs assessments against templates and reports results and violations. AWS Advisor, on the other hand, works at a higher level and provide guidance to provision resources following AWS best practices

Only standard EC2 reserved instances can be sold, not convertible type

A single NAT Gateway in each availability zone is enough. NAT Gateway is already redundant in nature, meaning, AWS already handles any failures that occur in your NAT Gateway in an availability zone.

RDS events only provide operational events such as DB instance events, DB parameter group events, DB security group events, and DB snapshot events. What we need in the scenario is to capture data-modifying events (INSERT, DELETE, UPDATE) which can be achieved thru native functions or stored procedures.

**\* RDS EVENTS DON'T DO CRUD OPERATIONS**

A Dedicated instance runs in a VPC on hardware that's dedicated to a single customer

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.

Amazon S3 forms the storage layer for Lake Formation. If you already use S3, you typically begin by registering existing S3 buckets that contain your data. Lake Formation creates new buckets for the data lake and import data into them. AWS always stores this data in your account, and only you have direct access to it.

Aurora can directly trigger Lambda. RDS doesn't acknowledge CRUD events

S3 object lock: (only works in versioned buckets)

- \* Retention period — Specifies a fixed period of time during which an object remains locked.

During this period, your object is WORM-protected and can't be overwritten or deleted. For more information, see Retention periods

- \* Legal hold — Provides the same protection as a retention period, but it has no expiration date. Instead, a legal hold remains in place until you explicitly remove it. Legal holds are independent from retention periods. For more information, see Legal holds

VPC are free by themselves

LDAP can be read by SSO

EBS max storage 64TB, S3 ONLY 5 TB (remember 3+2 =5 )

If not SAML then develop custom broker

Lambda@edge lets you run code closer to the users

RDS processes – Shows a summary of the resources used by the RDS management agent, diagnostics monitoring processes, and other AWS processes that are required to support RDS DB instances.

OS processes – Shows a summary of the kernel and system processes, which generally have minimal impact on performance.

CPU Utilization, Database Connections, and Freeable Memory are incorrect because these are just the regular items provided by Amazon RDS Metrics in CloudWatch. Remember that the scenario is asking for the Enhanced Monitoring metrics.

Only Global Accelerator does Anycast IP (one IP for one region)

SWF -> applications can pickup right where they left off

X-ray -> better than cloudwatch as can follow requests and where they go. Whereas the other just tracks API requests on AWS resources

RAID 0 makes non-persistent data go brrrrr

Application Load Balancers now support a slow start mode that allows you to add new targets without overwhelming them with a flood of requests

CloudTrail encrypts by default

CNAME.something.com

A.A.com

WAF can do geomatch conditions (not ALB)

Mgmt Events -> new resources used up etc

Data Events -> data level calls

EKSConnectorAgentRole only used initially for the first connection, no VPC endpoints besides S3/DynamoDB

Check if Aurora Serverless and Read Replicas are even mentioned/used. Aurora first attempts to start another instance in same AZ

ALB has access logs

Not Volume Gateway, only File Gateway can read NFS SMB

VPC endpoints are not inter-region

Gateway VPC endpoint > Interface VPC endpoint for simpler/less customization