A multi-national retail company has multiple business divisions, with each division having its own AWS account. The engineering team at the company would like to debug and trace data across these AWS accounts and visualize it in a centralized account.

As a Solutions Architect, which of the following solutions would you suggest for the given use-case?

ANS: X-ray

AWS X-Ray helps developers analyze and debug production, distributed applications, such as those built using a microservices architecture. With X-Ray, you can understand how your application and its underlying services are performing to identify and troubleshoot the root cause of performance issues and errors. X-Ray provides an end-to-end view of requests as they travel through your application, and shows a map of your application’s underlying components.

You can use X-Ray to collect data across AWS Accounts. The X-Ray agent can assume a role to publish data into an account different from the one in which it is running. This enables you to publish data from various components of your application into a central account.

A media company wants a low-latency way to distribute live sports results which are delivered via a proprietary application using UDP protocol.

As a solutions architect, which of the following solutions would you recommend such that it offers the BEST performance for this use case?

ANS: gloabal accelerator

A financial services company recently launched an initiative to improve the security of its AWS resources and it had enabled AWS Shield Advanced across multiple AWS accounts owned by the company. Upon analysis, the company has found that the costs incurred are much higher than expected.

Which of the following would you attribute as the underlying reason for the unexpectedly high costs for AWS Shield Advanced service?

ANS: Consolidate billing not enabled

The engineering team at an e-commerce company has been tasked with migrating to a serverless architecture. The team wants to focus on the key points of consideration when using Lambda as a backbone for this architecture.

As a Solutions Architect, which of the following options would you identify as correct for the given requirement? (Select three)

ANS:1. By default, Lambda functions always operate from an AWS-owned VPC and hence have access to any public internet address or public AWS APIs. Once a Lambda function is VPC-enabled, it will need a route through a NAT gateway in a public subnet to access public resources

2. Since Lambda functions can scale extremely quickly, its a good idea to deploy a CloudWatch Alarm that notifies your team when function metrics such as ConcurrentExecutions or Invocations exceeds the expected threshold

3. If you intend to reuse code in more than one Lambda function, you should consider creating a Lambda Layer for the reusable code

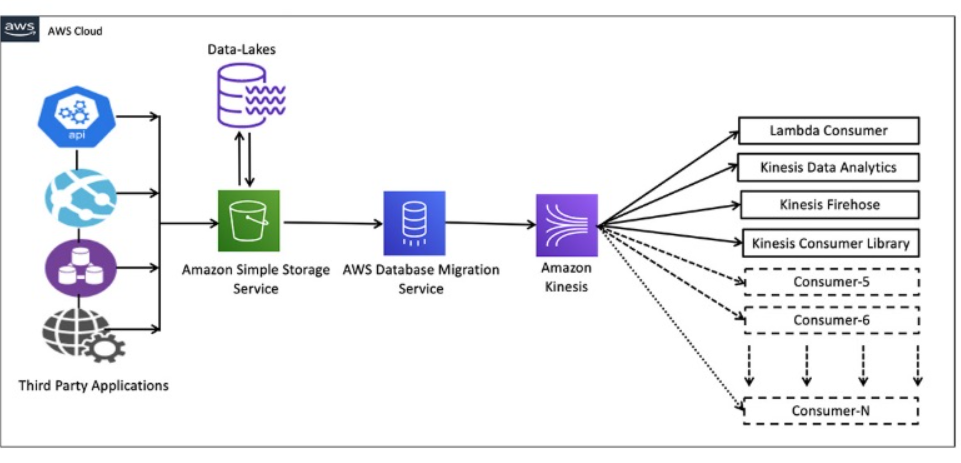
NOTE: The total unzipped size of the function and all layers can't exceed the unzipped deployment package size limit of 250 MB.  
----------------------------------------------------------------------------------------------------------

A Big Data analytics company writes data and log files in Amazon S3 buckets. The company now wants to stream the existing data files as well as any ongoing file updates from Amazon S3 to Amazon Kinesis Data Streams.

As a Solutions Architect, which of the following would you suggest as the fastest possible way of building a solution for this requirement?

ANS: Correct option:

Leverage AWS Database Migration Service (AWS DMS) as a bridge between Amazon S3 and Amazon Kinesis Data Streams

AWS DMS supports specifying Amazon S3 as the source and streaming services like Kinesis and Amazon Managed Streaming of Kafka (Amazon MSK) as the target. AWS DMS allows migration of full and change data capture (CDC) files to these services. AWS DMS performs this task out of box without any complex configuration or code development. You can also configure an AWS DMS replication instance to scale up or down depending on the workload. 

NOTE: 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. So this option is incorrect.

A media company wants to get out of the business of owning and maintaining its own IT infrastructure. As part of this digital transformation, the media company wants to archive about 5PB of data in its on-premises data center to durable long term storage.

As a solutions architect, what is your recommendation to migrate this data in the MOST cost-optimal way?

ANS: using snowball to se=3 and then using lifecycle policy to s3 glacier. YOU cannot directly copy data to s3 glacier from snowball

You can only use a launch template to provision capacity across multiple instance types using both On-Demand Instances and Spot Instances to achieve the desired scale, performance, and cost.

A company wants to migrate its on-premises databases to AWS Cloud. The CTO at the company wants a solution that can handle complex database configurations such as secondary indexes, foreign keys, and stored procedures.

ANS: database migration service(DMS) and Schema conversion tool(SCT)

Notes:

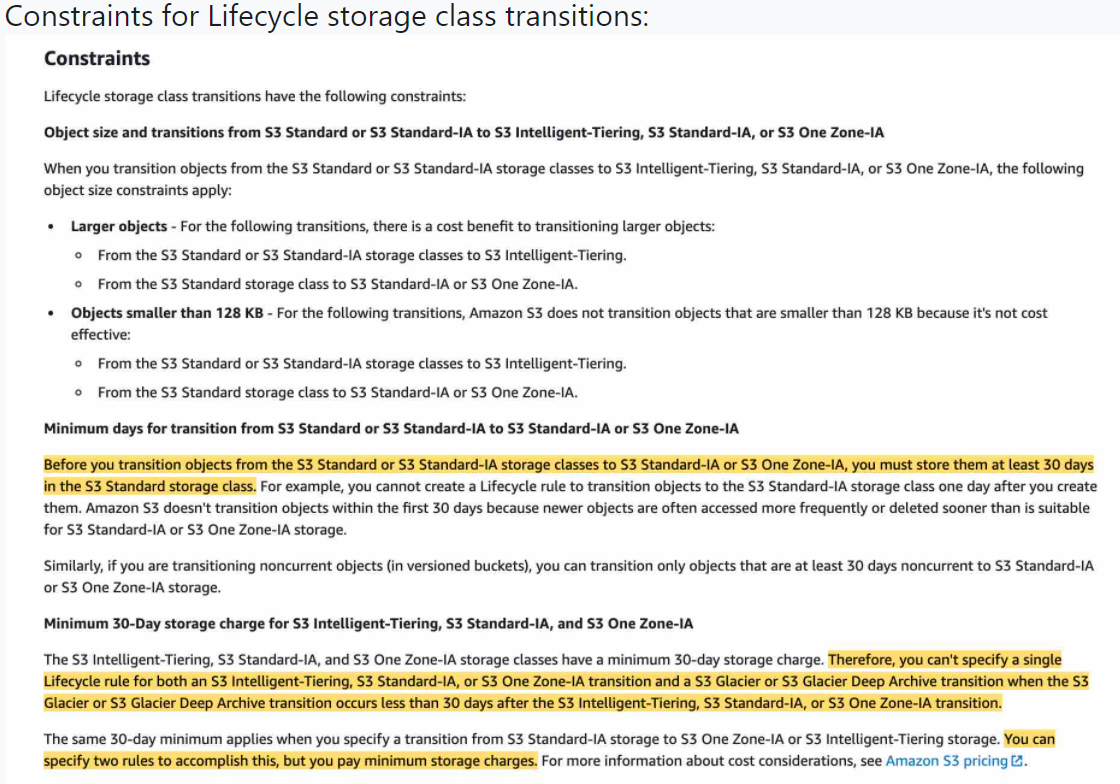
AWS Snowball Edge cannot be used for database migrations.

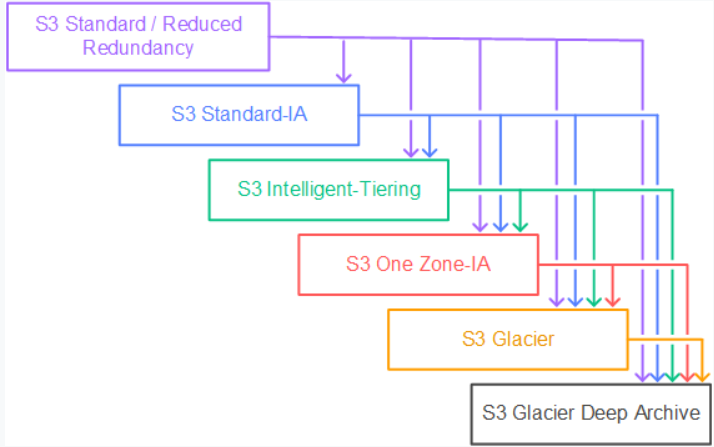
**Basic Schema Copy** : Basic Schema Copy will automatically create tables and primary keys in the target instance if the target does not already contain tables with the same names. Basic Schema Copy is great for doing a test migration, or when you are migrating databases heterogeneously e.g. Oracle to MySQL or SQL Server to Oracle. Basic Schema Copy will not migrate secondary indexes, foreign keys or stored procedures

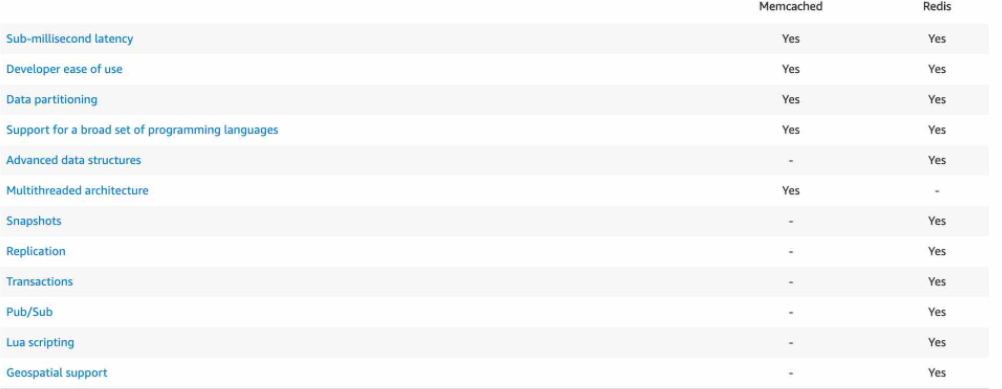
A medium-sized business has a taxi dispatch application deployed on an EC2 instance. Because of an unknown bug, the application causes the instance to freeze regularly. Then, the instance has to be manually restarted via the AWS management console.

ANS: Setup a CloudWatch alarm to monitor the health status of the instance. In case of an Instance Health Check failure, an EC2 Reboot CloudWatch Alarm Action can be used to reboot the instance

Using Amazon CloudWatch alarm actions, you can create alarms that automatically stop, terminate, reboot, or recover your EC2 instances. You can use the stop or terminate actions to help you save money when you no longer need an instance to be running.







A pharmaceutical company is considering moving to AWS Cloud to accelerate the research and development process. Most of the daily workflows would be centered around running batch jobs on EC2 instances with storage on EBS volumes. The CTO is concerned about meeting HIPAA compliance norms for sensitive data stored on EBS.

ANS: Correct options:

**Data at rest inside the volume is encrypted**

**Any snapshot created from the volume is encrypted**

**Data moving between the volume and the instance is encrypted**

**-------------------------------------------------------------------------------------------------------------------------**

A company has moved its business critical data to Amazon EFS file system which will be accessed by multiple EC2 instances.

As an AWS Certified Solutions Architect Associate, which of the following would you recommend to exercise access control such that only the permitted EC2 instances can read from the EFS file system? (Select three)

Correct options:

**Use VPC security groups to control the network traffic to and from your file system**

**Attach an IAM policy to your file system to control clients who can mount your file system with the required permissions**

**Use EFS Access Points to manage application access**

**-------------------------------------------------------------------------------------------------------------------------**

You have been hired as a Solutions Architect to advise a company on the various authentication/authorization mechanisms that AWS offers to authorize an API call within the API Gateway. The company would prefer a solution that offers built-in user management.

Which of the following solutions would you suggest as the best fit for the given use-case?

**Correct: Use Amazon Cognito User Pools** - A user pool is a user directory in Amazon Cognito. You can leverage Amazon Cognito User Pools to either provide built-in user management or integrate with external identity providers, such as Facebook, Twitter, Google+, and Amazon

Note: **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

A leading online gaming company is migrating its flagship application to AWS Cloud for delivering its online games to users across the world. The company would like to use a Network Load Balancer (NLB) to handle millions of requests per second. The engineering team has provisioned multiple instances in a public subnet and specified these instance IDs as the targets for the NLB.

As a solutions architect, can you help the engineering team understand the correct routing mechanism for these target instances?

ANS: If you specify targets using an instance ID, traffic is routed to instances using the primary private IP address specified in the primary network interface for the instance. The load balancer rewrites the destination IP address from the data packet before forwarding it to the target instance.

An engineering team wants to examine the feasibility of the user data feature of Amazon EC2 for an upcoming project.

Which of the following are true about the EC2 user data configuration? (Select two)

ANS: **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.

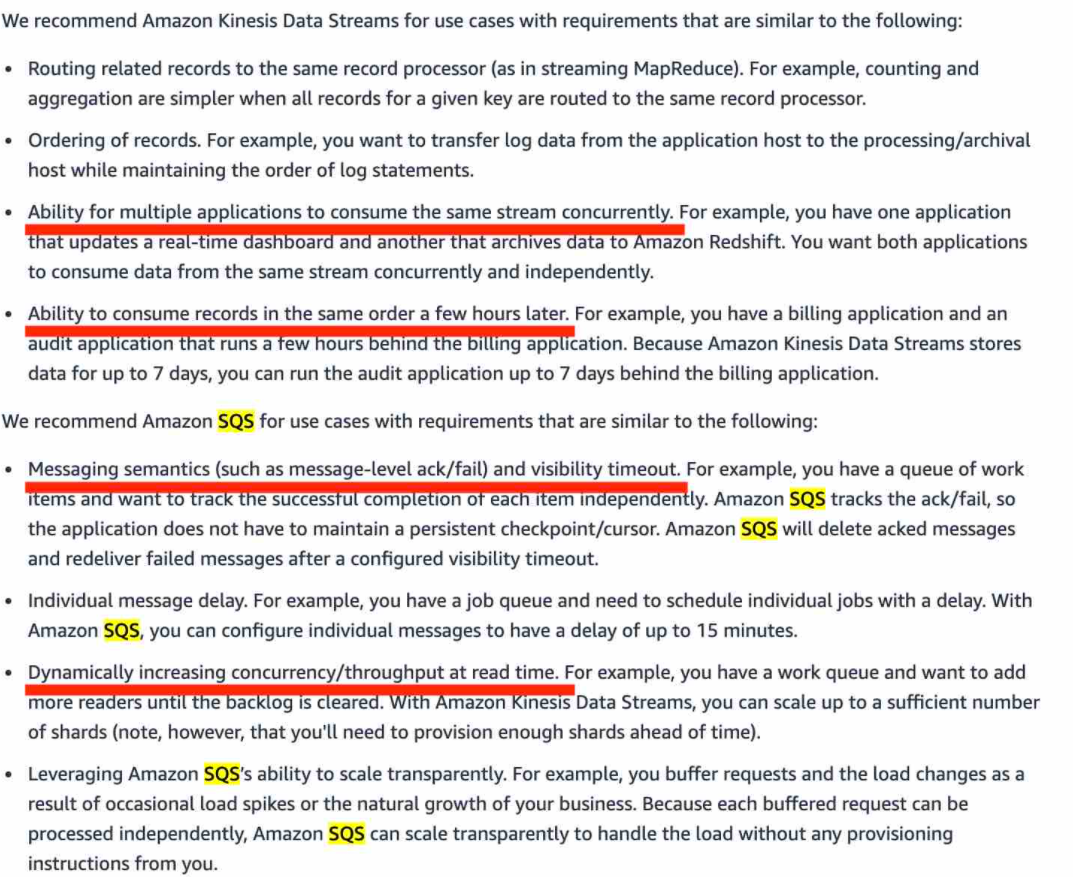
**By default, user data runs only during the boot cycle when you first launch an instance**

A big data analytics company is using Kinesis Data Streams (KDS) to process IoT data from the field devices of an agricultural sciences company. Multiple consumer applications are using the incoming data streams and the engineers have noticed a performance lag for the data delivery speed between producers and consumers of the data streams.

As a solutions architect, which of the following would you recommend for improving the performance for the given use-case?

**Use Enhanced Fanout feature of Kinesis Data Streams**

You should use enhanced fan-out if you have multiple consumers retrieving data from a stream in parallel. With enhanced fan-out developers can register stream consumers to use enhanced fan-out and receive their own 2MB/second pipe of read throughput per shard, and this throughput automatically scales with the number of shards in a stream.



The DevOps team at an IT company is provisioning a two-tier application in a VPC with a public subnet and a private subnet. The team wants to use either a NAT instance or a NAT gateway in the public subnet to enable instances in the private subnet to initiate outbound IPv4 traffic to the internet but needs some technical assistance in terms of the configuration options available for the NAT instance and the NAT gateway.

As a solutions architect, which of the following options would you identify as CORRECT? (Select three)

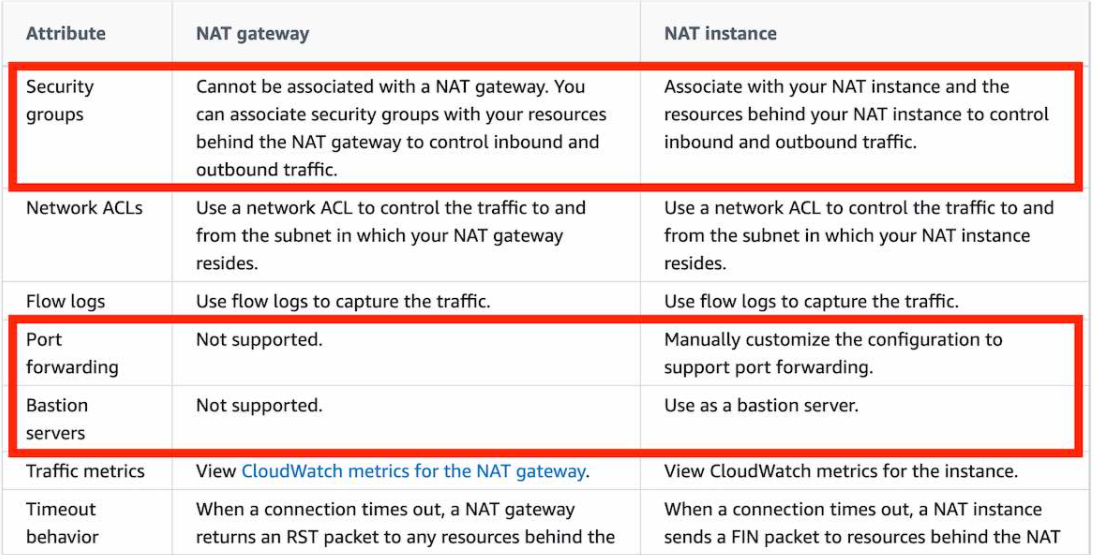
ANS: Correct options:

NAT instance can be used as a bastion server

Security Groups can be associated with a NAT instance

NAT instance supports port forwarding

NOTE: Port forwarding: Port forwarding, sometimes called port mapping, allows computers or services in private networks to connect over the internet with other public or private computers or services. Port forwarding achieves by creating an association called a map between a router’s public, wide area network (WAN) [**internet protocol (IP) address**](https://learn.g2.com/ip-address) and a private, local area network (LAN) IP address for a device on that private network.



An e-commerce company has copied 1 PB of data from its on-premises data center to an Amazon S3 bucket in the us-west-1 Region using an AWS Direct Connect link. The company now wants to copy the data to another S3 bucket in the us-east-1 Region. The on-premises data center does not allow the use of AWS Snowball.

As a Solutions Architect, which of the following would you recommend to accomplish this?

ANS: **Copy data from the source bucket to the destination bucket using the aws S3 sync command**

**Why not : Set up cross-Region replication to copy objects across S3 buckets in different Regions ? - As the data already exists in the source bucket, so you cannot use cross-Region replication because, by default, replication only supports copying new Amazon S3 objects after it is enabled.**

A retail company wants to rollout and test a blue-green deployment for its global application in the next 48 hours. Most of the customers use mobile phones which are prone to DNS caching. The company has only two days left for the annual Thanksgiving sale to commence.

As a Solutions Architect, which of the following options would you recommend to test the deployment on as many users as possible in the given time frame?

ANS: Blue/green deployment is a technique for releasing applications by shifting traffic between two identical environments running different versions of the application: "Blue" is the currently running version and "green" the new version

**Use AWS Global Accelerator to distribute a portion of traffic to a particular deployment** - AWS Global Accelerator uses endpoint weights to determine the proportion of traffic that is directed to endpoints in an endpoint group, and traffic dials to control the percentage of traffic that is directed to an endpoint group

NOTE: While relying on the DNS service is a great option for blue/green deployments, it may not fit use-cases that require a fast and controlled transition of the traffic.

The business analytics team at a company has been running ad-hoc queries on Oracle and PostgreSQL services on Amazon RDS to prepare daily reports for senior management. To facilitate the business analytics reporting, the engineering team now wants to continuously replicate this data and consolidate these databases into a petabyte-scale data warehouse by streaming data to Amazon Redshift.

As a solutions architect, which of the following would you recommend as the MOST resource-efficient solution that requires the LEAST amount of development time without the need to manage the underlying infrastructure?

ANS: **Use AWS Database Migration Service to replicate the data from the databases into Amazon Redshift**

**WHY NOT – kinesis data streams:**

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.

A Big Data analytics company wants to set up an AWS cloud architecture that throttles requests in case of sudden traffic spikes. The company is looking for AWS services that can be used for buffering or throttling to handle such traffic variations.

Which of the following services can be used to support this requirement?

ANS: **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.

A company has grown from a small startup to an enterprise employing over 1000 people. As the team size has grown, the company has recently observed some strange behavior, with S3 buckets **settings** being changed regularly.

How can you figure out what's happening without restricting the rights of the users?

ANS: **Use CloudTrail to analyze API calls**

**Why not Use S3 access logs to analyze user access using Athena ?-**

Amazon S3 server access logging **provides detailed records for the requests that are made to a bucke**t. Server access logs are useful for many applications. For example, access log information can be useful in security and access audits. It can also help you learn about your customer base and understand your Amazon S3 bill. S3 access logs would not provide us the necessary information, so it's not the correct choice for this use-case.

The engineering team at an e-commerce company wants to migrate from SQS Standard queues to FIFO queues with batching.

As a solutions architect, which of the following steps would you have in the migration checklist? (Select three)

ANS: Correct options:

**Delete the existing standard queue and recreate it as a FIFO queue :** You can't convert an existing standard queue into a FIFO queue. To make the move, you must either create a new FIFO queue for your application or delete your existing standard queue and recreate it as a FIFO queue.

**Make sure that the name of the FIFO queue ends with the .fifo suffix**

**Make sure that the throughput for the target FIFO queue does not exceed 3,000 messages per second**

A company has noticed that its application performance has deteriorated after a new Auto Scaling group was deployed a few days back. Upon investigation, the team found out that the Launch Configuration selected for the Auto Scaling group is using the incorrect instance type that is not optimized to handle the application workflow.

As a solutions architect, what would you recommend to provide a long term resolution for this issue?

* 

**Modify the launch configuration to use the correct instance type and continue to use the existing Auto Scaling group**

* 

**No need to modify the launch configuration. Just modify the Auto Scaling group to use more number of existing instance types. More instances may offset the loss of performance**

* 

**No need to modify the launch configuration. Just modify the Auto Scaling group to use the correct instance type**

* 

**Create a new launch configuration to use the correct instance type. Modify the Auto Scaling group to use this new launch configuration. Delete the old launch configuration as it is no longer needed**

**(Correct)**

**Explanation :** A launch configuration is an instance configuration template that an Auto Scaling group uses to launch EC2 instances. When you create a launch configuration, you specify information for the instances. Include the ID of the Amazon Machine Image (AMI), the instance type, a key pair, one or more security groups, and a block device mapping.

It is not possible to modify a launch configuration once it is created. The correct option is to create a new launch configuration to use the correct instance type.

An e-commerce company is planning to migrate their two-tier application from on-premises infrastructure to AWS Cloud. As the engineering team at the company is new to the AWS Cloud, they are planning to use the Amazon VPC console wizard to set up the networking configuration for the two-tier application having public web servers and private database servers.

Can you spot the configuration that is NOT supported by the Amazon VPC console wizard?

ANS: **VPC with a public subnet only and AWS Site-to-Site VPN access**" is NOT supported by the Amazon VPC console wizard.

VPC supports:

VPC with a single public subnet

VPC with public and private subnets (NAT)

VPC with public and private subnets and AWS Site-to-Site VPN access

VPC with a private subnet only and AWS Site-to-Site VPN access

A financial services firm uses a high-frequency trading system and wants to write the log files into Amazon S3. The system will also read these log files in parallel on a near real-time basis. The engineering team wants to address any data discrepancies that might arise when the trading system overwrites an existing log file and then tries to read that specific log file.

Which of the following options BEST describes the capabilities of Amazon S3 relevant to this scenario?

ANS: A process replaces an existing object and immediately tries to read it. Amazon S3 always returns the latest version of the object

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.

After a successful write of a new object or an overwrite of an existing object, any subsequent read request immediately receives the latest version of the object. S3 **also provides strong consistency for list operations**, so after a write, you can immediately perform a listing of the objects in a bucket with any changes reflected.

To summarize, all S3 GET, PUT, and LIST operations, as well as operations that change object tags, ACLs, or metadata, are strongly consistent.

A data analytics company measures what the consumers watch and what advertising they’re exposed to. This real-time data is ingested into its on-premises data center and subsequently, the daily data feed is compressed into a single file and uploaded on Amazon S3 for backup. The typical compressed file size is around 2 GB.

Which of the following is the fastest way to upload the daily compressed file into S3?

ANS: Upload the compressed file using multipart upload with S3 transfer acceleration

A major bank is using SQS to migrate several core banking applications to the cloud to ensure high availability and cost efficiency while simplifying administrative complexity and overhead. The development team at the bank expects a peak rate of about 1000 messages per second to be processed via SQS. It is important that the messages are processed in order.

Which of the following options can be used to implement this system?

ANS: Use Amazon SQS FIFO queue in batch mode of 4 messages per operation to process the messages at the peak rate

By default, FIFO queues support up to 300 messages per second (300 send, receive, or delete operations per second). When you batch 10 messages per operation (maximum), FIFO queues can support up to 3,000 messages per second. Therefore you need to process 4 messages per operation so that the FIFO queue can support up to 1200 messages per second, which is well within the peak rate.

AWS Lambda supports runtimes for the following languages:

C#/.NET

Go

Java

Node.js

Python

Ruby

Does not support: C, PHP, R

A financial services company uses Amazon GuardDuty for analyzing its AWS account metadata to meet the compliance guidelines. However, the company has now decided to stop using GuardDuty service. All the existing findings have to be deleted and cannot persist anywhere on AWS Cloud.

Which of the following techniques will help the company meet this requirement?

ANS: **Disable the service in the general setting**s - Disabling the service will delete all remaining data, including your findings and configurations before relinquishing the service permissions and resetting the service. So, this is the correct option for our use case.

NOTE: **Suspend** the service in the general settings - You can stop Amazon GuardDuty from analyzing your data sources at any time by choosing to suspend the service in the general settings. This will immediately stop the service from analyzing data, but **does not delete your existing findings or configurations**.

CloudFront offers a multi-tier cache in the form of regional edge caches that improve latency. However, there are certain content types that bypass the regional edge cache, and go directly to the origin.

Which of the following content types skip the regional edge cache? (Select two)

ANS: **Dynamic content, as determined at request time** (cache-behavior configured to forward all headers)

**Proxy methods PUT/POST/PATCH/OPTIONS/DELETE** go directly to the origin from the POPs and do not proxy through the regional edge caches. So this option is also correct.

The engineering team at a Spanish professional football club has built a notification system for its website using Amazon SNS notifications which are then handled by a Lambda function for end-user delivery. During the off-season, the notification systems need to handle about 100 requests per second. During the peak football season, the rate touches about 5000 requests per second and it is noticed that a significant number of the notifications are not being delivered to the end-users on the website.

As a solutions architect, which of the following would you suggest as the BEST possible solution to this issue?

ANS: Amazon SNS message deliveries to AWS Lambda have crossed the account concurrency quota for Lambda, so the team needs to contact AWS support to raise the account limit

A large IT company wants to federate its workforce into AWS accounts and business applications.

Which of the following AWS services can help build a solution for this requirement? (Select two)

ANS: Use AWS Single Sign-On (SSO)

Use AWS Identity and Access Management (IAM)

An ivy-league university is assisting NASA to find potential landing sites for exploration vehicles of unmanned missions to our neighboring planets. The university uses High Performance Computing (HPC) driven application architecture to identify these landing sites.

Which of the following EC2 instance topologies should this application be deployed on?

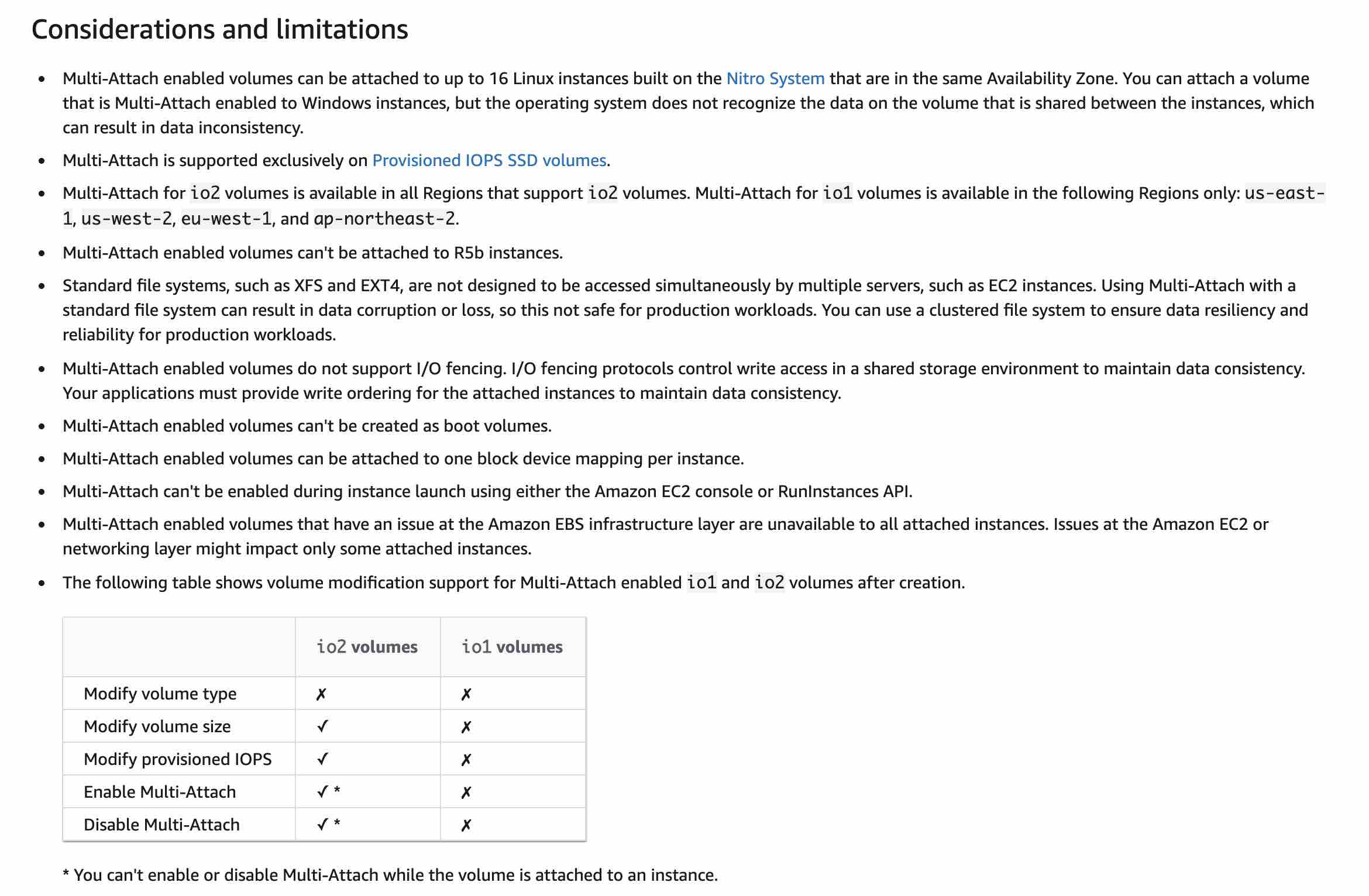
ANS: The EC2 instances should be deployed in a cluster placement group so that the underlying workload can benefit from low network latency and high network throughput

A company wants some EBS volumes with maximum possible Provisioned IOPS (PIOPS) to support high-performance database workloads on EC2 instances. The company also wants some EBS volumes that can be attached to multiple EC2 instances in the same Availability Zone.

As an AWS Certified Solutions Architect Associate, which of the following options would you identify as correct for the given requirements? (Select two)

ANS: Use io2 **Block Express** volumes on **Nitro-based EC2** instances to achieve a maximum Provisioned IOPS of 256,000

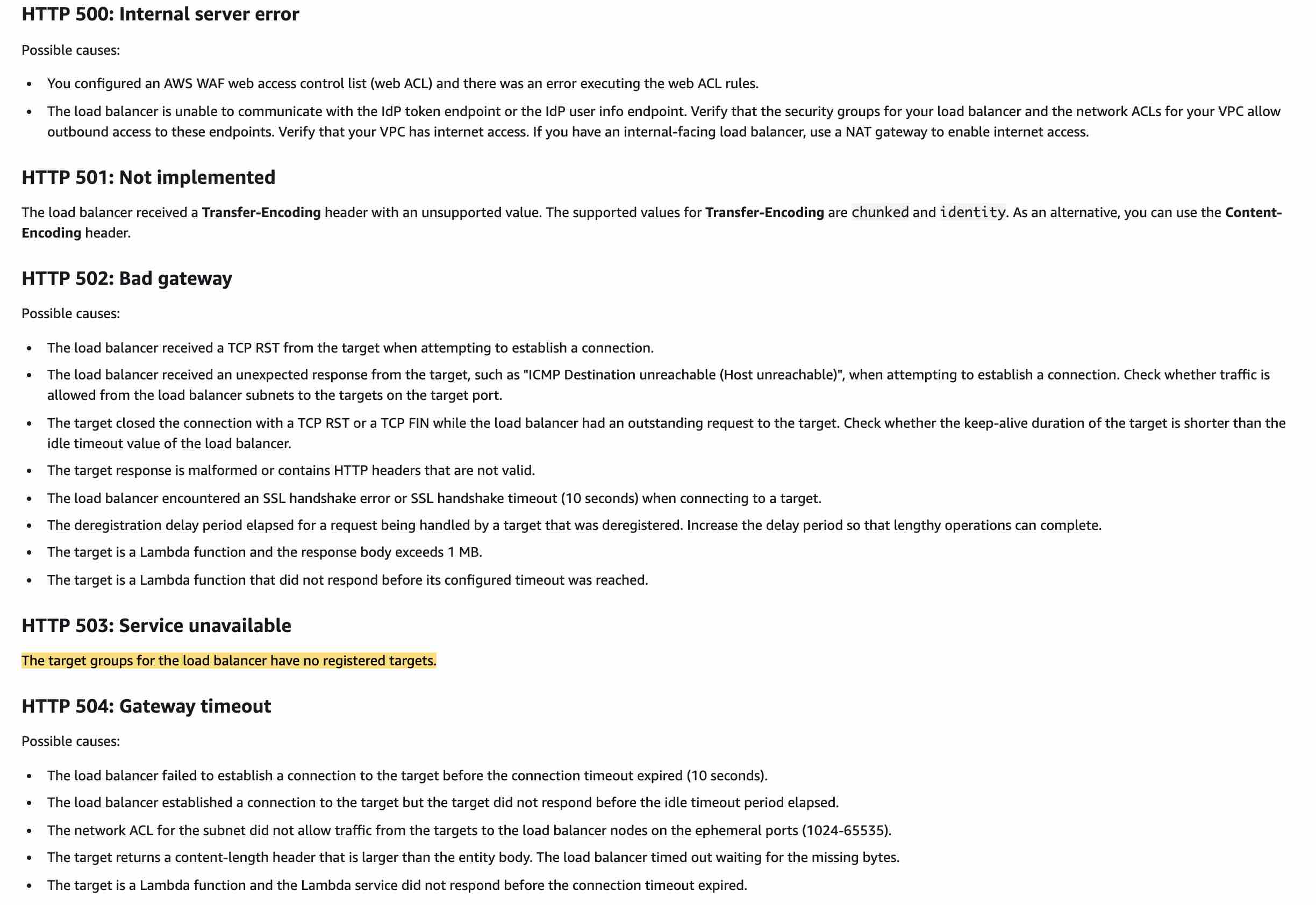
Use io1/io2 volumes to enable Multi-Attach on Nitro-based EC2 instances



A developer has created a new Application Load Balancer but has not registered any targets with the target groups. Which of the following errors would be generated by the Load Balancer?

ANS: HTTP 503: Service unavailable

The Load Balancer generates the HTTP 503: Service unavailable error when the target groups for the load balancer have no registered targets.



The product team at a startup has figured out a market need to support both stateful and stateless client-server communications via the APIs developed using its platform. You have been hired by the startup as a solutions architect to build a solution to fulfill this market need using AWS API Gateway.

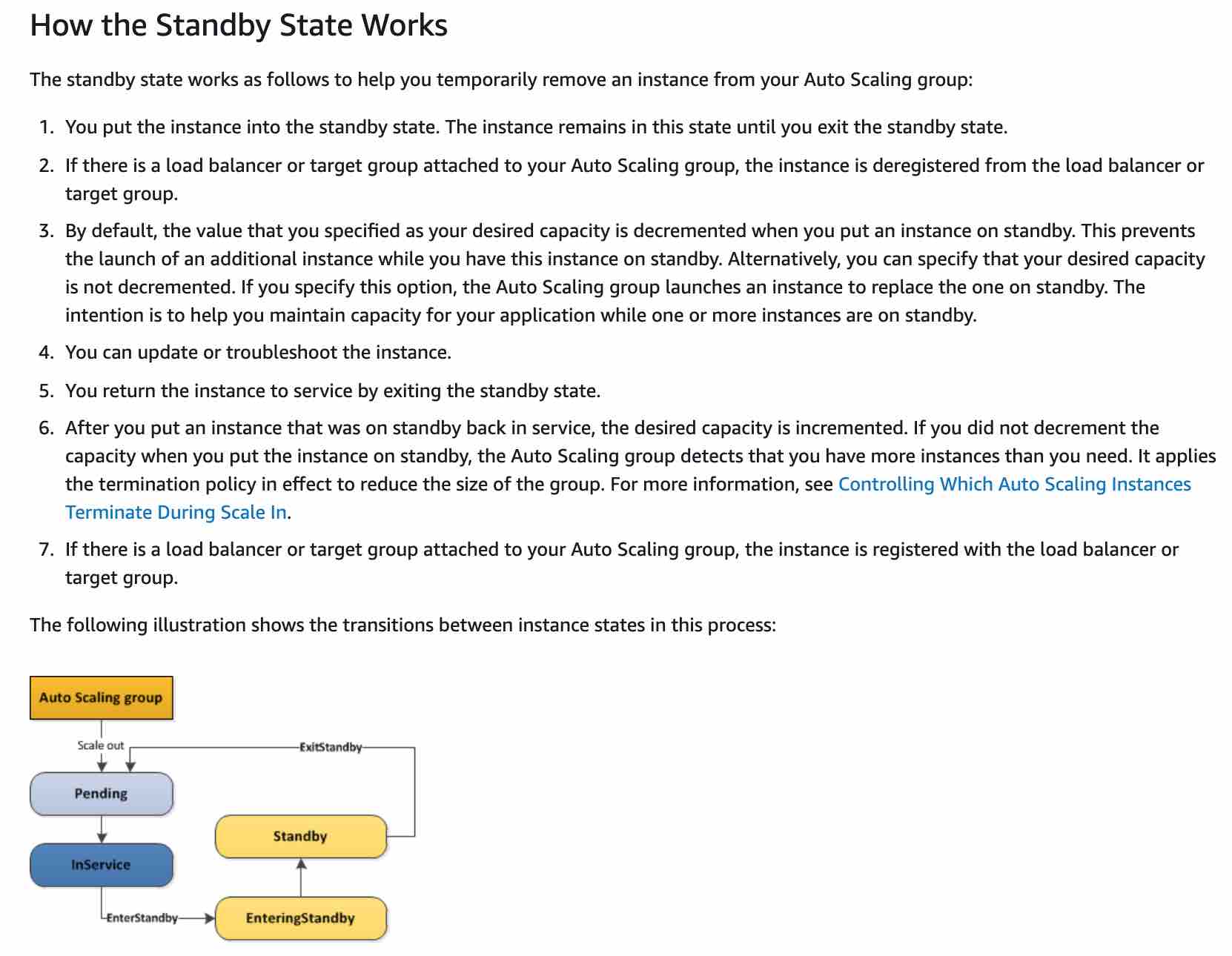
Which of the following would you identify as correct?

ANS: API Gateway creates RESTful APIs that enable stateless client-server communication and API Gateway also creates WebSocket APIs that adhere to the WebSocket protocol, which enables stateful, full-duplex communication between client and server

The DevOps team at an e-commerce company wants to perform some maintenance work on a specific EC2 instance that is part of an Auto Scaling group using a step scaling policy. The team is facing a maintenance challenge - every time the team deploys a maintenance patch, the instance health check status shows as out of service for a few minutes. This causes the Auto Scaling group to provision another replacement instance immediately.

As a solutions architect, which are the MOST time/resource efficient steps that you would recommend so that the maintenance work can be completed at the earliest? (Select two)

ANS: Put the instance into the Standby state and then update the instance by applying the maintenance patch. Once the instance is ready, you can exit the Standby state and then return the instance to service



**Suspend the ReplaceUnhealthy process type for the Auto Scaling group and apply the maintenance patch to the instance. Once the instance is ready, you can manually set the instance's health status back to healthy and activate the ReplaceUnhealthy process type again** –

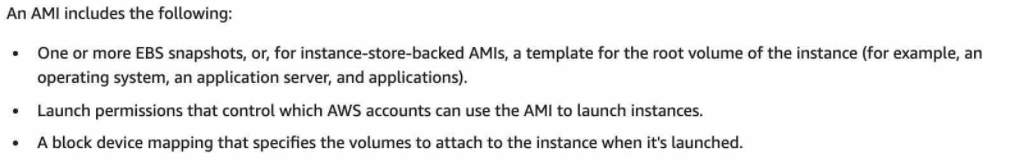
The ReplaceUnhealthy process terminates instances that are marked as unhealthy and then creates new instances to replace them. Amazon EC2 Auto Scaling stops replacing instances that are marked as unhealthy. Instances that fail EC2 or Elastic Load Balancing health checks are still marked as unhealthy. As soon as you resume the ReplaceUnhealthly process, Amazon EC2 Auto Scaling replaces instances that were marked unhealthy while this process was suspended.

The solo founder at a tech startup has just created a brand new AWS account. The founder has provisioned an EC2 instance 1A which is running in region A. Later, he takes a snapshot of the instance 1A and then creates a new AMI in region A from this snapshot. This AMI is then copied into another region B. The founder provisions an instance 1B in region B using this new AMI in region B.

At this point in time, what entities exist in region B?

ANS: 1 EC2 instance, 1 AMI and 1 snapshot exist in region B

An Amazon Machine Image (AMI) provides the information required to launch an instance. You must specify an AMI when you launch an instance. When the new AMI is copied from region A into region B, it automatically creates a snapshot in region B because AMIs are based on the underlying snapshots. Further, an instance is created from this AMI in region B. Hence, we have 1 EC2 instance, 1 AMI and 1 snapshot in region B.



A research group needs a fleet of EC2 instances for a specialized task that must deliver high random I/O performance. Each instance in the fleet would have access to a dataset that is replicated across the instances. Because of the resilient application architecture, the specialized task would continue to be processed even if any instance goes down, as the underlying application architecture would ensure the replacement instance has access to the required dataset.

Which of the following options is the MOST cost-optimal and resource-efficient solution to build this fleet of EC2 instances?

ANS: **Use Instance Store based EC2 instances**

As Instance Store based volumes provide high random I/O performance at low cost (as the storage is part of the instance's usage cost) and the resilient architecture can adjust for the loss of any instance, therefore you should use Instance Store based EC2 instances for this use-case.

Incorrect options:

Use EBS based EC2 instances - EBS based volumes would need to use Provisioned IOPS (io1) as the storage type and that would incur additional costs. As we are looking for the most cost-optimal solution, this option is ruled out.

Use EC2 instances with EFS mount points - Using EFS implies that extra resources would have to be provisioned. As we are looking for the most resource-efficient solution, this option is also ruled out.

Use EC2 instances with access to S3 based storage - Using EC2 instances with access to S3 based storage does not deliver high random I/O performance, this option is just added as a distractor.

A leading video streaming service delivers billions of hours of content from Amazon S3 to customers around the world. Amazon S3 also serves as the data lake for its big data analytics solution. The data lake has a staging zone where intermediary query results are kept only for 24 hours. These results are also heavily referenced by other parts of the analytics pipeline.

Which of the following is the MOST cost-effective strategy for storing this intermediary query data?

ANS: **Store the intermediary query results in S3 Standard storage class**

WHY NOT - Store the intermediary query results in S3 One Zone-Infrequent Access storage class?

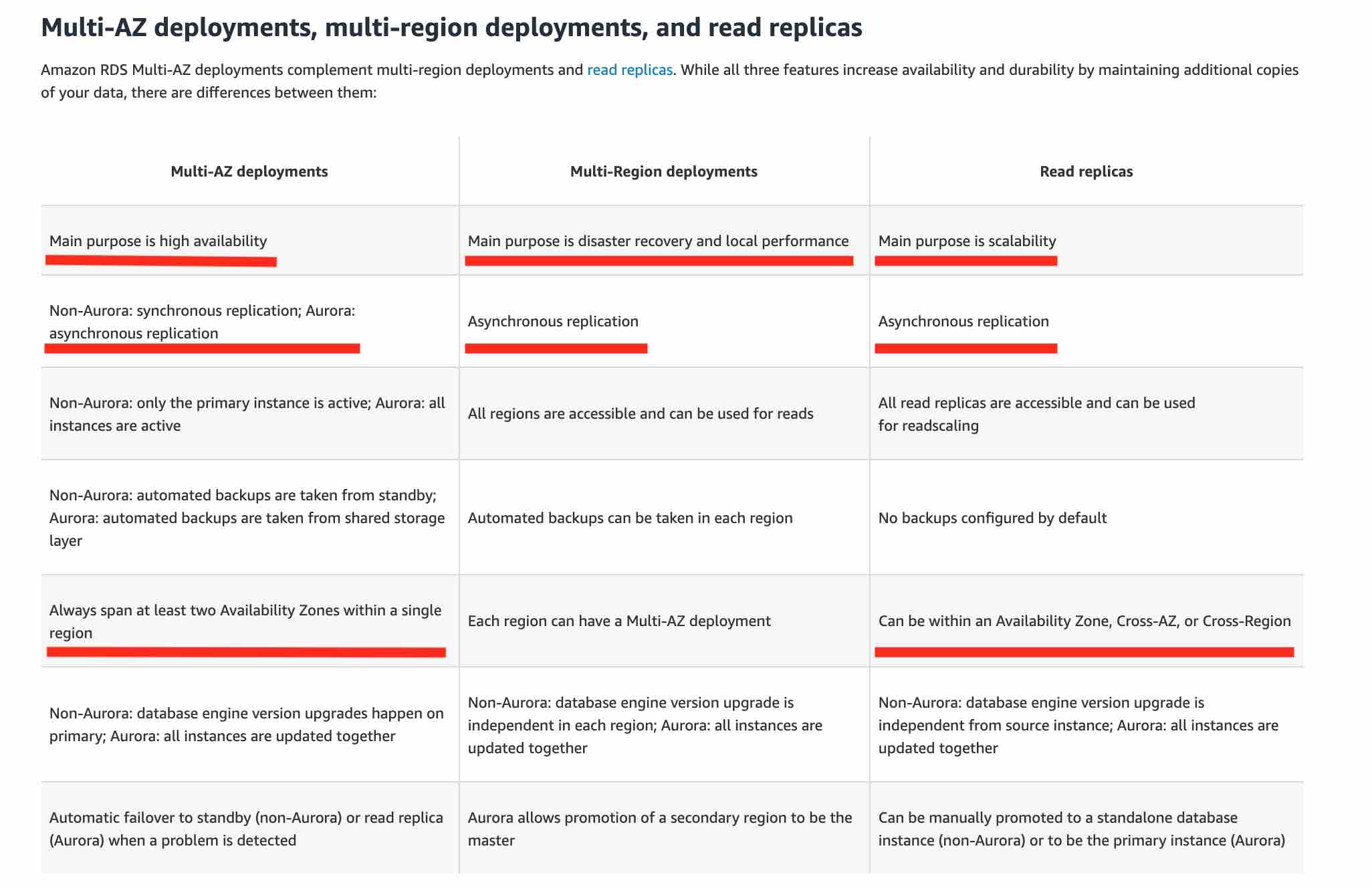
S3 Standard-IA and S3 One Zone-IA have a minimum storage duration charge of 30 days (so instead of 24 hours, you end up paying for 30 days). S3 Standard-IA and S3 One Zone-IA **also have retrieval charges** (as the results are heavily referenced by other parts of the analytics pipeline, so the retrieval costs would be pretty high). Therefore, these storage classes are not cost optimal for the given use-case.

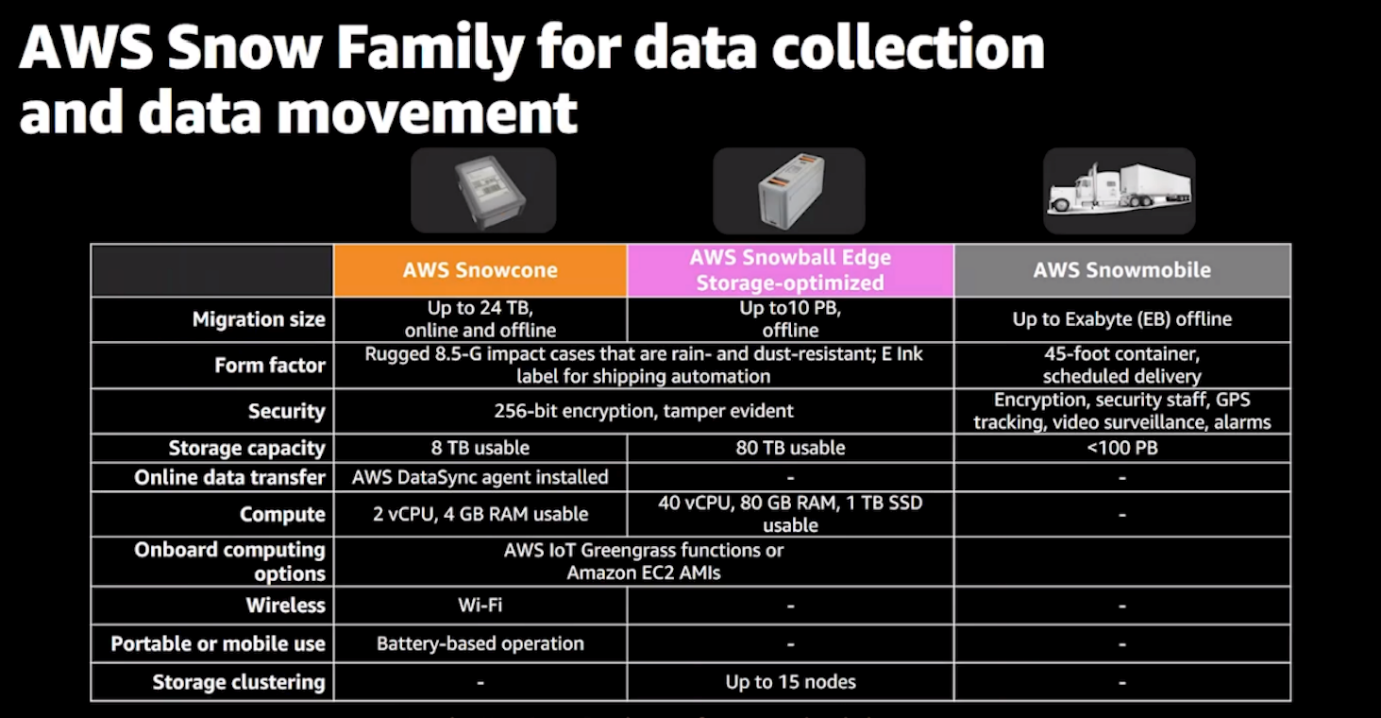
The sourcing team at the US headquarters of a global e-commerce company is preparing a spreadsheet of the new product catalog. The spreadsheet is saved on an EFS file system created in us-east-1 region. The sourcing team counterparts from other AWS regions such as Asia Pacific and Europe also want to collaborate on this spreadsheet.

As a solutions architect, what is your recommendation to enable this collaboration with the LEAST amount of operational overhead?

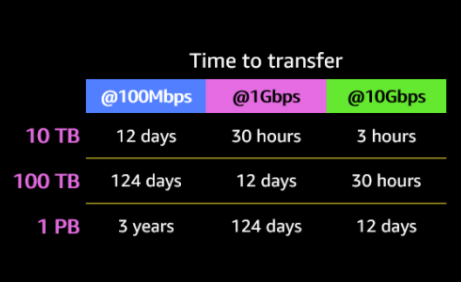
ANS: The spreadsheet on the EFS file system can be accessed in other AWS regions by using an inter-region VPC peering connection

WHY NOT - **The spreadsheet will have to be copied into EFS file systems of other AWS regions as EFS is a regional service and it does not allow access from other AWS regions** - Creating copies of the spreadsheet into EFS file systems of other AWS regions would mean no collaboration would be possible between the teams. In this case, each team would work on "its own file" instead of a single file accessed and updated by all teams. Hence this option is incorrect.





Time to tranfer over internet :



A video analytics organization has been acquired by a leading media company. The analytics organization has 10 independent applications with an on-premises data footprint of about 70TB for each application. The CTO of the media company has set a timeline of two weeks to carry out the data migration from on-premises data center to AWS Cloud and establish connectivity.

Which of the following are the MOST cost-effective options for completing the data transfer and establishing connectivity? (Select two)

ANS: **Order 10 Snowball Edge Storage Optimized devices to complete the one-time data transfer**

**Setup Site-to-Site VPN to establish connectivity between the on-premises data center and AWS Cloud- can use VPN to transfer upto 100 TB of data over a 1 GBPS connection in 12 days .** Each AWS Site-to-Site VPN connection has two tunnels and each tunnel supports a maximum throughput of up to 1.25 Gbps.

A company uses Amazon S3 buckets for storing sensitive customer data. The company has defined different retention periods for different objects present in the Amazon S3 buckets, based on the compliance requirements. But, the retention rules do not seem to work as expected.

Which of the following options represent a valid configuration for setting up retention periods for objects in Amazon S3 buckets? (Select two)

ANS: **When you apply a retention period to an object version explicitly, you specify a Retain Until Date for the object version** - You can place a retention period on an object version either explicitly or through a bucket default setting. When you apply a retention period to an object version explicitly, you specify a Retain Until Date for the object version. Amazon S3 stores the Retain Until Date setting in the object version's metadata and protects the object version until the retention period expires.

**Different versions of a single object can have different retention modes and periods** - Like all other Object Lock settings, retention periods apply to individual object versions. Different versions of a single object can have different retention modes and periods.

INcorrecct Options:

**When you use bucket default settings, you specify a Retain Until Date for the object version** - When you use bucket default settings, you don't specify a Retain Until Date. Instead, you specify a duration, in either days or years, for which every object version placed in the bucket should be protected.

**The bucket default settings will override any explicit retention mode or period you request on an object version** - If your request to place an object version in a bucket contains an explicit retention mode and period, those settings override any bucket default settings for that object version.

A gaming company is looking at improving the availability and performance of its global flagship application which utilizes UDP protocol and needs to support fast regional failover in case an AWS Region goes down. The company wants to continue using its own custom DNS service.

Which of the following AWS services represents the best solution for this use-case?

ANS: **AWS Global Accelerator**

Global Accelerator improves performance for a wide range of applications over TCP or UDP by proxying packets at the edge to applications running in one or more AWS Regions. Global Accelerator is a good fit for non-HTTP use cases, such as gaming (UDP), IoT (MQTT), or Voice over IP, as well as for HTTP use cases that specifically require static IP addresses or deterministic, fast regional failover

INCORRECT: AWS Elastic Load Balancing (ELB)

Both of the services, ELB and Global Accelerator solve the challenge of routing user requests to healthy application endpoints. 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.

If you have workloads that cater to a global client base, AWS recommends that you use AWS Global Accelerator. If you have workloads hosted in a single AWS Region and used by clients in and around the same Region, you can use an Application Load Balancer or Network Load Balancer to manage your resources.

A social photo-sharing company uses Amazon S3 to store the images uploaded by the users. These images are kept encrypted in S3 by using AWS-KMS and the company manages its own Customer Master Key (CMK) for encryption. A member of the DevOps team accidentally deleted the CMK a day ago, thereby rendering the user's photo data unrecoverable. You have been contacted by the company to consult them on possible solutions to this crisis.

As a solutions architect, which of the following steps would you recommend to solve this issue?

ANS: **As the CMK was deleted a day ago, it must be in the 'pending deletion' status and hence you can just cancel the CMK deletion and recover the key**

Deleting a customer master key (CMK) in AWS Key Management Service (AWS KMS) is destructive and potentially dangerous. Therefore, AWS KMS enforces a waiting period. To delete a CMK in AWS KMS you schedule key deletion. You can set the waiting period from a minimum of 7 days up to a maximum of 30 days. The default waiting period is 30 days. During the waiting period, the CMK status and key state is Pending deletion. To recover the CMK, you can cancel key deletion before the waiting period ends. After the waiting period ends you cannot cancel key deletion, and AWS KMS deletes the CMK.

A company has moved its business critical data to Amazon EFS file system which will be accessed by multiple EC2 instances.

As an AWS Certified Solutions Architect Associate, which of the following would you recommend to exercise access control such that only the permitted EC2 instances can read from the EFS file system? (Select three)

ANS: **Use VPC security groups to control the network traffic to and from your file system**

**Attach an IAM policy to your file system to control clients who can mount your file system with the required permissions**

**Use EFS Access Points to manage application access**

Files and directories in an Amazon EFS file system support standard Unix-style read, write, and execute permissions based on the user ID and group IDs. When an NFS client mounts an EFS file system without using an access point, the user ID and group ID provided by the client is trusted. You can use EFS access points to override user ID and group IDs used by the NFS client. When users attempt to access files and directories, Amazon EFS checks their user IDs and group IDs to verify that each user has permission to access the objects

A file-hosting service uses Amazon S3 under the hood to power its storage offerings. Currently all the customer files are uploaded directly under a single S3 bucket. The engineering team has started seeing scalability issues where customer file uploads have started failing during the peak access hours with more than 5000 requests per second.

Which of the following is the MOST resource efficient and cost-optimal way of addressing this issue?

ANS: **Change the application architecture to create customer-specific custom prefixes within the single bucket and then upload the daily files into those prefixed locations**

S3 can achieve at least 3,500 PUT/COPY/POST/DELETE or 5,500 GET/HEAD requests per second per prefix in a bucket.

There are no limits to the number of prefixes in a bucket. You can increase your read or write performance by parallelizing reads. For example, if you create 10 prefixes in an Amazon S3 bucket to parallelize reads, you could scale your read performance to 55,000 read requests per second

A media company runs a photo-sharing web application that is accessed across three different countries. The application is deployed on several Amazon EC2 instances running behind an Application Load Balancer. With new government regulations, the company has been asked to block access from two countries and allow access only from the home country of the company.

Which configuration should be used to meet this changed requirement?

ANS: **Configure AWS WAF on the Application Load Balancer in a VPC**

You can use AWS WAF with your Application Load Balancer to allow or block requests based on the rules in a web access control list (web ACL). Geographic (Geo) Match Conditions in AWS WAF allows you to use AWS WAF to restrict application access based on the geographic location of your viewers. With geo match conditions you can choose the countries from which AWS WAF should allow access.

**Incorrect**:

**Use Geo Restriction feature of Amazon CloudFront in a VPC** - Geo Restriction feature of CloudFront helps in restricting traffic based on the user's geographic location. But, CloudFront works from edge locations and doesn't belong to a VPC. Hence, this option itself is incorrect and given only as a distractor.

**Configure the security group on the Application Load Balancer**

**Configure the security group for the EC2 instances**

Security Groups cannot restrict access based on the user's geographic location.

A gaming company uses Amazon Aurora as its primary database service. The company has now deployed 5 multi-AZ read replicas to increase the read throughput and for use as failover target. The replicas have been assigned the following failover priority tiers and corresponding instance sizes are given in parentheses: tier-1 (16TB), tier-1 (32TB), tier-10 (16TB), tier-15 (16TB), tier-15 (32TB).

In the event of a failover, Amazon Aurora will promote which of the following read replicas?

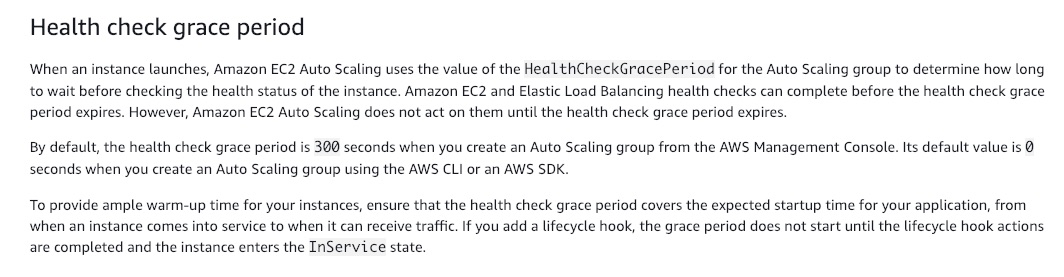
ANS: **Tier-1 (32TB)**

For Amazon Aurora, each Read Replica is associated with a priority tier (0-15). In the event of a failover, Amazon Aurora will promote the Read Replica that has the highest priority (the lowest numbered tier). If two or more Aurora Replicas share the same priority, then Amazon RDS promotes the replica that is largest in size. If two or more Aurora Replicas share the same priority and size, then Amazon Aurora promotes an arbitrary replica in the same promotion tier.

The engineering team at a logistics company has noticed that the Auto Scaling group (ASG) is not terminating an unhealthy Amazon EC2 instance.

As a Solutions Architect, which of the following options would you suggest to troubleshoot the issue? (Select three)

ANS: **The health check grace period for the instance has not expired** - Amazon EC2 Auto Scaling doesn't terminate an instance that came into service based on EC2 status checks and ELB health checks until the health check grace period expires.



**The instance maybe in Impaired status** - Amazon EC2 Auto Scaling does not immediately terminate instances with an Impaired status. Instead, Amazon EC2 Auto Scaling waits a few minutes for the instance to recover. Amazon EC2 Auto Scaling might also delay or not terminate instances that fail to report data for status checks. This usually happens when there is insufficient data for the status check metrics in Amazon CloudWatch.

**The instance has failed the ELB health check status** - By default, Amazon EC2 Auto Scaling doesn't use the results of ELB health checks to determine an instance's health status when the group's health check configuration is set to EC2. As a result, Amazon EC2 Auto Scaling doesn't terminate instances that fail ELB health checks. If an instance's status is OutofService on the ELB console, but the instance's status is Healthy on the Amazon EC2 Auto Scaling console, confirm that the health check type is set to ELB.

Incorrect options:

**The EC2 instance could be a spot instance type, which cannot be terminated by ASG** - This is an incorrect statement. Amazon EC2 Auto Scaling terminates Spot instances when capacity is no longer available or the Spot price exceeds your maximum price.

**A user might have updated the configuration of ASG and increased the minimum number of instances forcing ASG to keep all instances alive** - This statement is incorrect. If the configuration is updated and ASG needs more number of instances, ASG will launch new, healthy instances and does not keep unhealthy ones alive.

**A custom health check might have failed. ASG does not terminate instances that are set unhealthy by custom checks** - This statement is incorrect. You can define custom health checks in Amazon EC2 Auto Scaling. When a custom health check determines that an instance is unhealthy, the check manually triggers SetInstanceHealth and then sets the instance's state to Unhealthy. Amazon EC2 Auto Scaling then terminates the unhealthy instance.

A financial services company wants a single log processing model for all the log files (consisting of system logs, application logs, database logs, etc) that can be processed in a serverless fashion and then durably stored for downstream analytics. The company wants to use an AWS managed service that automatically scales to match the throughput of the log data and requires no ongoing administration.

As a solutions architect, which of the following AWS services would you recommend solving this problem?

ANS: Correct option:

**Kinesis Data Firehose**

Amazon Kinesis Data Firehose is the easiest way to reliably load streaming data into data lakes, data stores, and analytics tools. It can capture, transform, and load streaming data into Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk, enabling near real-time analytics with existing business intelligence tools and dashboards you’re already using today. It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration. Therefore, this is the correct option.

Incorrect options:

**Kinesis Data Streams** - 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). As it requires manual administration of shards, it's not the correct choice for the given use-case.

**Amazon EMR** - Amazon EMR uses Hadoop, an open-source framework, to distribute your data and processing across a resizable cluster of Amazon EC2 instances.

Using an EMR cluster would imply managing the underlying infrastructure so it’s ruled out.

**AWS Lambda** - AWS Lambda lets you run code without provisioning or managing servers. It cannot be used for production-grade serverless log analytics.

You have a team of developers in your company, and you would like to ensure they can quickly experiment with AWS Managed Policies by attaching them to their accounts, but you would like to prevent them from doing an escalation of privileges, by granting themselves the AdministratorAccess managed policy. How should you proceed?

ANS: **For each developer, define an IAM permission boundary that will restrict the managed policies they can attach to themselves**

WHY NOT - Put the developers into an IAM group, and then define an IAM permission boundary on the group that will restrict the managed policies they can attach to themselves ?

**- IAM permission boundary can only be applied to roles or users, not IAM groups.** Hence this option is incorrect.

A Machine Learning research group uses a proprietary computer vision application hosted on an EC2 instance. Every time the instance needs to be stopped and started again, the application takes about 3 minutes to start as some auxiliary software programs need to be executed so that the application can function. The research group would like to minimize the application boostrap time whenever the system needs to be **stopped** and then started at a later point in time.

As a solutions architect, which of the following solutions would you recommend for this use-case?

ANS: **Use EC2 Instance Hibernate**

A company is developing a healthcare application that cannot afford any downtime for database write operations. The company has hired you as an AWS Certified Solutions Architect Associate to build a solution using Amazon Aurora.

Which of the following options would you recommend?

ANS: S**et up an Aurora multi-master DB cluster**

In a multi-master cluster, all DB instances can perform write operations. There isn't any failover when a writer DB instance becomes unavailable, because another writer DB instance is immediately available to take over the work of the failed instance. AWS refers to this type of availability as continuous availability, to distinguish it from the high availability (with brief downtime during failover) offered by a single-master cluster.

Incorrect options:

**Set up an Aurora serverless DB cluster**

**Set up an Aurora provisioned DB cluster**

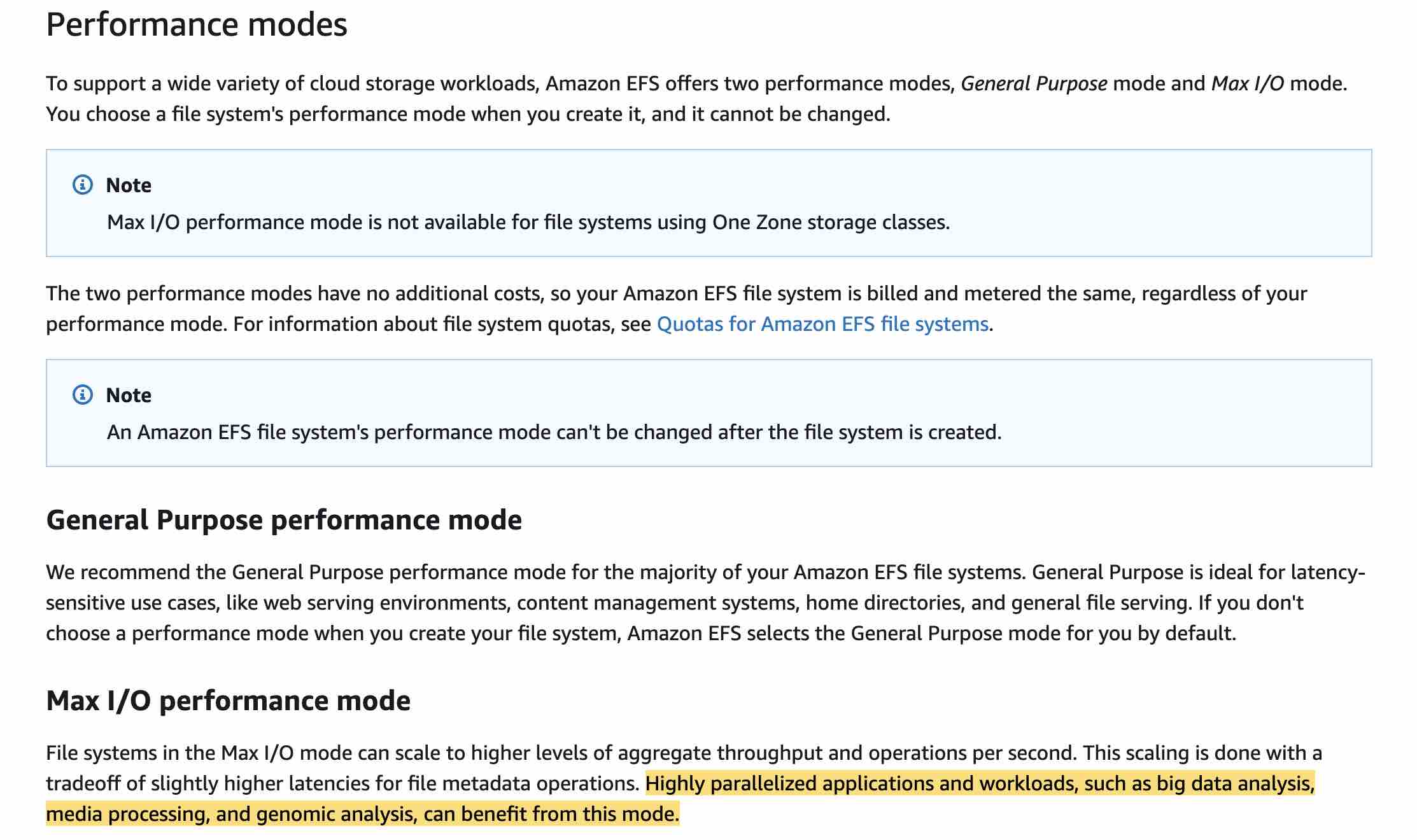
**Set up an Aurora Global Database cluster**

These three options represent Aurora single-master clusters. In a single-master cluster, a single DB instance performs all write operations and any other DB instances are read-only. If the writer DB instance becomes unavailable, a failover mechanism promotes one of the read-only instances to be the new writer. As there is **a brief downtime during this failover**, so these three options are incorrect for the given use case.

An analytics company wants to improve the performance of its big data processing workflows running on Amazon EFS. Which of the following performance modes should be used for EFS to address this requirement?

ANS: **Max I/O**

Max I/O performance mode is used to scale to higher levels of aggregate throughput and operations per second. This scaling is done with a tradeoff of slightly higher latencies for file metadata operations. Highly parallelized applications and workloads, such as big data analysis, media processing, and genomic analysis, can benefit from this mode.



Incorrect options:

**Provisioned Throughput**

**Bursting Throughput**

These two options have been added as distractors as these refer to the **throughput mode of EFS and not the performance mode**. There are two throughput modes to choose from for your file system, Bursting Throughput and Provisioned Throughput. With **Bursting Throughput mode, throughput on Amazon EFS scales as the size of your file system in the standard storage class grows**. **With Provisioned Throughput mode, you can instantly provision the throughput of your file system (in MiB/s) independent of the amount of data stored.**

An HTTP application is deployed on an Auto Scaling Group, is accessible from an Application Load Balancer that provides HTTPS termination, and accesses a PostgreSQL database managed by RDS.

How should you configure the security groups? (Select three)

ANS: **The security group of RDS should have an inbound rule from the security group of the EC2 instances in the ASG on port 5432**

**The security group of the EC2 instances should have an inbound rule from the security group of the ALB on port 80**

**The security group of the ALB should have an inbound rule from anywhere on port 443**

PostgreSQL port = 5432 HTTP port = 80 HTTPS port = 443

The traffic goes like this : The client sends an HTTPS request to ALB on port 443. This is handled by the rule - The security group of the ALB should have an inbound rule from anywhere on port 443. The ALB then forwards the request to one of the EC2 instances. This is handled by the rule - The security group of the EC2 instances should have an inbound rule from the security group of the ALB on port 80. The EC2 instance further accesses the PostgreSQL database managed by RDS on port 5432. This is handled by the rule - The security group of RDS should have an inbound rule from the security group of the EC2 instances in the ASG on port 5432.

A retail company wants to share sensitive accounting data that is stored in an Amazon RDS DB instance with an external auditor. The auditor has its own AWS account and needs its own copy of the database.

Which of the following would you recommend to securely share the database with the auditor?

ANS: **Create an encrypted snapshot of the database, share the snapshot, and allow access to the AWS Key Management Service (AWS KMS) encryption key**

You can share the AWS Key Management Service (AWS KMS) customer master key (CMK) that was used to encrypt the snapshot with any accounts that you want to be able to access the snapshot. You can share AWS KMS CMKs with another AWS account by adding the other account to the AWS KMS key policy.

Making an encrypted snapshot of the database will give the auditor a copy of the database, as required for the given use case.

Incorrect options:

**Create a snapshot of the database in Amazon S3 and assign an IAM role to the auditor to grant access to the object in that bucket** - RDS stores the DB snapshots in the Amazon S3 bucket belonging to the same AWS region where the RDS instance is located. RDS stores these on your behalf and you do not have direct access to these snapshots in S3, so it's not possible to grant access to the snapshot objects in S3.

**Export the database contents to text files, store the files in Amazon S3, and create a new IAM user for the auditor with access to that bucket** - This solution is feasible though not optimal. It requires a lot of unnecessary work and is difficult to audit when such bulk data is exported into text files.

**Set up a read replica of the database and configure IAM standard database authentication to grant the auditor access** - Read Replicas make it easy to elastically scale out beyond the capacity constraints of a single DB instance for read-heavy database workloads. Creating Read Replicas for audit purposes is overkill. Also, the question mentions that the auditor needs to have their own copy of the database, which is not possible with replicas.

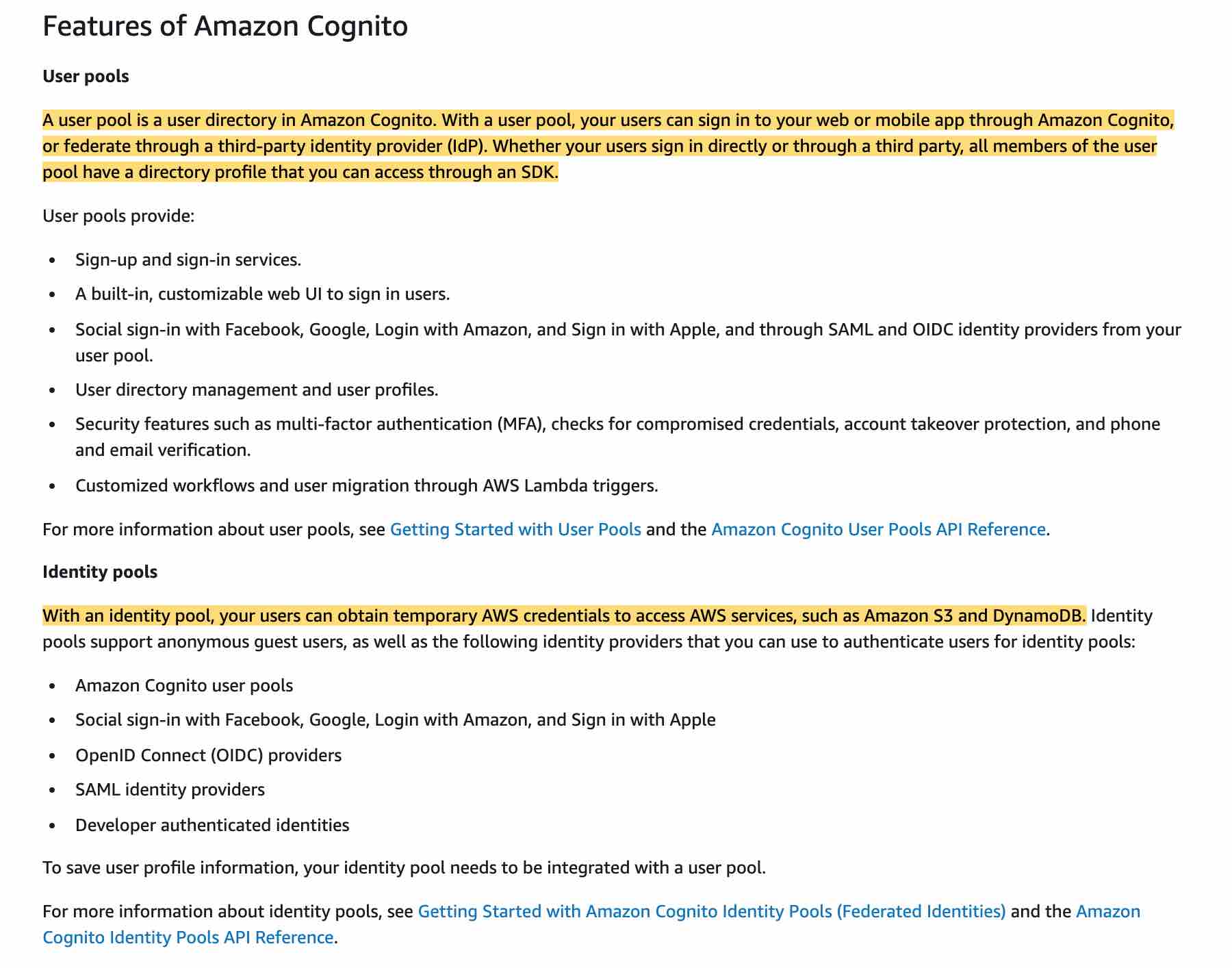
A social media application is hosted on an EC2 server fleet running behind an Application Load Balancer. The application traffic is fronted by a CloudFront distribution. The engineering team wants to decouple the user authentication process for the application, so that the application servers can just focus on the business logic.

As a Solutions Architect, which of the following solutions would you recommend to the development team so that it requires minimal development effort?

ANS: **Use Cognito Authentication via Cognito User Pools for your Application Load Balancer**

Application Load Balancer can be used to securely authenticate users for accessing your applications. This enables you to offload the work of authenticating users to your load balancer so that your applications can focus on their business logic. You can use Cognito User Pools to authenticate users through well-known social IdPs, such as Amazon, Facebook, or Google, through the user pools supported by Amazon Cognito or through corporate identities, using SAML, LDAP, or Microsoft AD, through the user pools supported by Amazon Cognito. You configure user authentication by creating an authenticate action for one or more listener rules.

Exam Alert:

Please review the following note to understand the differences between Cognito User Pools and Cognito Identity Pools:  

Incorrect Options:

**Use Cognito Authentication via Cognito Identity Pools for your Application Load Balance**r - **There is no such thing as using Cognito Authentication via Cognito Identity Pools for managing user authentication** for the application. Application-specific user authentication can be provided via Cognito User Pools. Amazon **Cognito identity pools provide temporary AWS credentials for users who are guests** (unauthenticated) and for users who have been authenticated and received a token.

**Use Cognito Authentication via Cognito User Pools for your CloudFront distribution** - You cannot directly integrate Cognito User Pools with CloudFront distribution as you have to create a separate Lambda@Edge function to accomplish the authentication via Cognito User Pools. This involves additional development effort, so this option is not the best fit for the given use-case.

What does this IAM policy do?

**{**

**"Version": "2012-10-17",**

**"Statement": [**

**{**

**"Sid": "Mystery Policy",**

**"Action": [**

**"ec2:RunInstances"**

**],**

**"Effect": "Allow",**

**"Resource": "\*",**

**"Condition": {**

**"StringEquals": {**

**"aws:RequestedRegion": "eu-west-1"**

**}**

**}**

**}**

**]**

**}**

ANS: **It allows running EC2 instances only in the eu-west-1 region, and the API call can be made from anywhere in the world**

You can use the aws:RequestedRegion key to compare the AWS Region that was called in the request with the Region that you specify in the policy. You can use this global condition key to control which Regions can be requested.

aws:RequestedRegion represents the target of the API call. So in this example, we can only launch EC2 instances in eu-west-1, and we can do this API call from anywhere.

The aws:SourceIP in this condition always represents the IP of the caller of the API. That is very helpful if you want to restrict access to certain AWS API for example from the public IP of your on-premises infrastructure.

An IT company is working on client engagement to build a real-time data analytics tool for the Internet of Things (IoT) data. The IoT data is funneled into Kinesis Data Streams which further acts as the source of a delivery stream for Kinesis Firehose. The engineering team has now configured a Kinesis Agent to send IoT data from another set of devices to the same Firehose delivery stream. They noticed that data is not reaching Firehose as expected.

As a solutions architect, which of the following options would you attribute as the MOST plausible root cause behind this issue?

ANS: **Kinesis Agent cannot write to a Kinesis Firehose for which the delivery stream source is already set as Kinesis Data Streams**

When a Kinesis data stream is configured as the source of a Firehose delivery stream, Firehose’s PutRecord and PutRecordBatch operations are disabled and Kinesis Agent cannot write to Firehose delivery stream directly. Data needs to be added to the Kinesis data stream through the Kinesis Data Streams PutRecord and PutRecords operations instead. Therefore, this option is correct.

A company has recently launched a new mobile gaming application that the users are adopting rapidly. The company uses RDS MySQL as the database. The engineering team wants an urgent solution to this issue where the rapidly increasing workload might exceed the available database storage.

As a solutions architect, which of the following solutions would you recommend so that it requires minimum development and systems administration effort to address this requirement?

ANS: **Enable storage auto-scaling for RDS MySQL**

If your workload is unpredictable, you can enable storage autoscaling for an Amazon RDS DB instance. With storage autoscaling enabled, when Amazon RDS detects that you are running out of free database space it automatically scales up your storage. Amazon RDS starts a storage modification for an autoscaling-enabled DB instance when these factors apply:

Free available space is less than 10 percent of the allocated storage.

The low-storage condition lasts at least five minutes.

At least six hours have passed since the last storage modification.

The maximum storage threshold is the limit that you set for autoscaling the DB instance. You can't set the maximum storage threshold for autoscaling-enabled instances to a value greater than the maximum allocated storage.

You are establishing a monitoring solution for desktop systems, that will be sending telemetry data into AWS every 1 minute. Data for each system must be processed in order, independently, and you would like to scale the number of consumers to be possibly equal to the number of desktop systems that are being monitored.

What do you recommend?

ANS: **Use an SQS FIFO queue, and make sure the telemetry data is sent with a Group ID attribute representing the value of the Desktop ID**

use an SQS FIFO queue. If we don't specify a GroupID, then all the messages are in absolute order, but we can only have 1 consumer at most. To allow for multiple consumers to read data for each Desktop application, and to scale the number of consumers, we should use the "Group ID" attribute. So this is the correct option.

WHY NOT- **Use a Kinesis Data Stream, and send the telemetry data with a Partition ID that uses the value of the Desktop ID**

A Kinesis Data Stream would work and would give us the data for each desktop application within shards, but we can only have as many consumers as shards in Kinesis (which is in practice, much less than the number of producers).

Amazon EC2 Auto Scaling needs to terminate an instance from Availability Zone (AZ) us-east-1a as it has the most number of instances amongst the AZs being used currently. There are 4 instances in the AZ us-east-1a like so: Instance A has the oldest launch template, Instance B has the oldest launch configuration, Instance C has the newest launch configuration and Instance D is closest to the next billing hour.

Which of the following instances would be terminated per the default termination policy?

ANS: **Instance B**

Per the default termination policy, the first priority is given to any allocation strategy for On-Demand vs Spot instances. As no such information has been provided for the given use-case, so this criterion can be ignored. The next priority is to consider any instance with the oldest launch template unless there is an instance that uses a launch configuration. So this rules out Instance A. Next, you need to consider any instance which has the oldest launch configuration. This implies Instance B will be selected for termination and Instance C will also be ruled out as it has the newest launch configuration. Instance D, which is closest to the next billing hour, is not selected as this criterion is last in the order of priority.

An IT company has built a solution wherein a Redshift cluster writes data to an Amazon S3 bucket belonging to a different AWS account. However, it is found that the files created in the S3 bucket using the UNLOAD command from the Redshift cluster are not even accessible to the S3 bucket owner.

What could be the reason for this denial of permission for the bucket owner?

ANS: **By default, an S3 object is owned by the AWS account that uploaded it. So the S3 bucket owner will not implicitly have access to the objects written by Redshift cluster** - By default, an S3 object is owned by the AWS account that uploaded it. This is true even when the bucket is owned by another account. Because the Amazon Redshift data files from the UNLOAD command were put into your bucket by another account, you (the bucket owner) don't have default permission to access those files.

To get access to the data files, an AWS Identity and Access Management (IAM) role with cross-account permissions must run the UNLOAD command again. Follow these steps to set up the Amazon Redshift cluster with cross-account permissions to the bucket:

1. From the account of the S3 bucket, create an IAM role (Bucket Role) with permissions to the bucket.
2. From the account of the Amazon Redshift cluster, create another IAM role (Cluster Role) with permissions to assume the Bucket Role.
3. Update the Bucket Role to grant bucket access and create a trust relationship with the Cluster Role.
4. From the Amazon Redshift cluster, run the UNLOAD command using the Cluster Role and Bucket Role.

This solution doesn't apply to Amazon Redshift clusters or S3 buckets that use server-side encryption with AWS Key Management Service (AWS KMS).

A freelance developer has built a Python based web application. The developer would like to upload his code to AWS Cloud and have AWS handle the deployment automatically. He also wants access to the underlying operating system for further enhancements.

As a solutions architect, which of the following AWS services would you recommend for this use-case?

ANS: AWS Elastic Beanstalk

you retain full control over the AWS resources powering your application and can access the underlying resources at any time. There is no additional charge for Elastic Beanstalk - you pay only for the AWS resources needed to store and run your applications.

A data analytics company is using SQS queues for decoupling the various processes of an application workflow. The company wants to postpone the delivery of certain messages to the queue by one minute while all other messages need to be delivered immediately to the queue.

As a solutions architect, which of the following solutions would you suggest to the company?

ANS: **Use message timers to postpone the delivery of certain messages to the queue by one minute**

You can use message timers to set an initial invisibility period for a message added to a queue. So, if you send a message with a 60-second timer, the message isn't visible to consumers for its first 60 seconds in the queue. The default (minimum) delay for a message is 0 seconds. The maximum is 15 minutes.

INCORRECT:

**Use delay queues to postpone the delivery of certain messages to the queue by one minute** - Delay queues let you postpone the delivery of all new messages to a queue for several seconds, for example, when your consumer application needs additional time to process messages

A gaming company uses Application Load Balancers (ALBs) in front of Amazon EC2 instances for different services and microservices. The architecture has now become complex with too many ALBs in multiple AWS Regions. Security updates, firewall configurations, and traffic routing logic have become complex with too many IP addresses and configurations.

The company is looking at an easy and effective way to bring down the number of IP addresses allowed by the firewall and easily manage the entire network infrastructure. Which of these options represents an appropriate solution for this requirement?

ANS: **Launch AWS Global Accelerator and create endpoints for all the Regions. Register the ALBs of each Region to the corresponding endpoints**

 When the internet is congested, Global Accelerator’s automatic routing optimizations will help keep your packet loss, jitter, and latency consistently low.

With Global Accelerator, you are provided two global static customer-facing IPs to simplify traffic management. On the back end, add or remove your AWS application origins, such as Network Load Balancers, Application Load Balancers, Elastic IPs, and EC2 Instances, without making user-facing changes. To mitigate endpoint failure, Global Accelerator automatically re-routes your traffic to your nearest healthy available endpoint.

A retail company has connected its on-premises data centre to the AWS Cloud via AWS Direct Connect. The company wants to be able to resolve DNS queries for any resources in the on-premises network from the AWS VPC and also resolve any DNS queries for resources in the AWS VPC from the on-premises network.

As a solutions architect, which of the following solutions can be combined to address the given use case? (Select two)

ANS: **Create an inbound endpoint on Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint**

**Create an outbound endpoint on Route 53 Resolver and then Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint**

By default, Route 53 Resolver automatically answers DNS queries for local VPC domain names for EC2 instances. You can integrate DNS resolution between Resolver and DNS resolvers on your on-premises network by configuring forwarding rules.

To resolve any DNS queries for resources in the AWS VPC from the on-premises network, you can create an inbound endpoint on Route 53 Resolver and then DNS resolvers on the on-premises network can forward DNS queries to Route 53 Resolver via this endpoint.

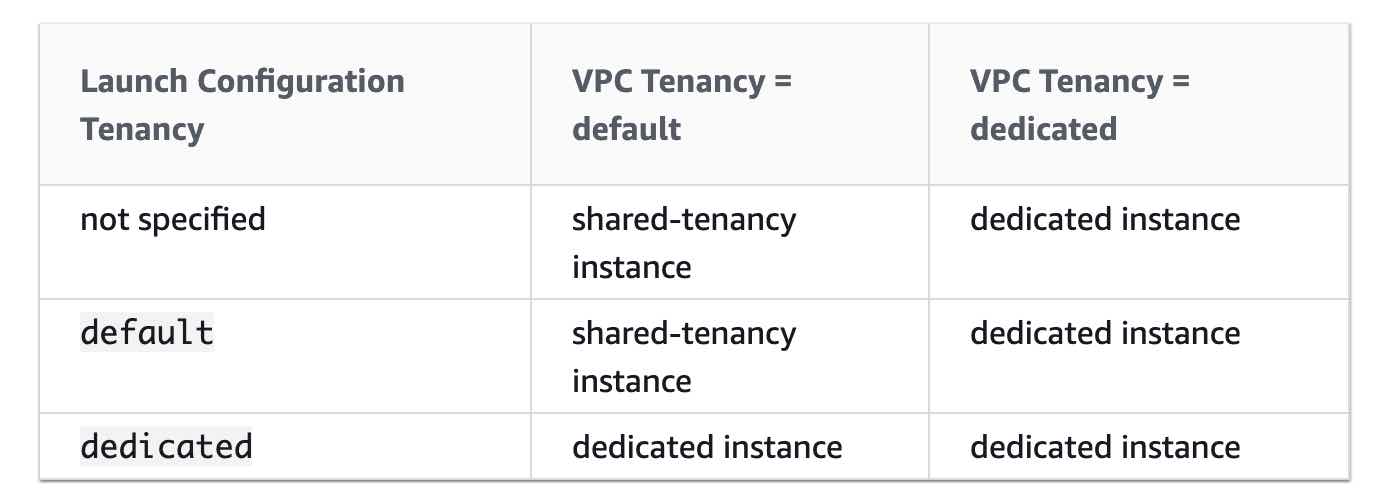
To resolve DNS queries for any resources in the on-premises network from the AWS VPC, you can create an outbound endpoint on Route 53 Resolver and then Route 53 Resolver can conditionally forward queries to resolvers on the on-premises network via this endpoint. To conditionally forward queries, you need to create Resolver rules that specify the domain names for the DNS queries that you want to forward (such as example.com) and the IP addresses of the DNS resolvers on the on-premises network that you want to forward the queries to.

A video conferencing application is hosted on a fleet of EC2 instances which are part of an Auto Scaling group (ASG). The ASG uses a Launch Configuration (LC1) with "dedicated" instance placement tenancy but the VPC (V1) used by the Launch Configuration LC1 has the instance tenancy set to default. Later the DevOps team creates a new Launch Configuration (LC2) with "default" instance placement tenancy but the VPC (V2) used by the Launch Configuration LC2 has the instance tenancy set to dedicated.

Which of the following is correct regarding the instances launched via Launch Configuration LC1 and Launch Configuration LC2?

ANS: **The instances launched by both Launch Configuration LC1 and Launch Configuration LC2 will have dedicated instance tenancy**

When you create a launch configuration, the default value for the instance placement tenancy is null and the instance tenancy is controlled by the tenancy attribute of the VPC. If you set the Launch Configuration Tenancy to default and the VPC Tenancy is set to dedicated, then the instances have dedicated tenancy. If you set the Launch Configuration Tenancy to dedicated and the VPC Tenancy is set to default, then again the instances have dedicated tenancy.

Launch Configuration Tenancy vs VPC Tenancy 

An IT company is looking to move its on-premises infrastructure to AWS Cloud. The company has a portfolio of applications with a few of them using server bound licenses that are valid for the next year. To utilize the licenses, the CTO wants to use dedicated hosts for a one year term and then migrate the given instances to default tenancy thereafter.

As a solutions architect, which of the following options would you identify as CORRECT for changing the tenancy of an instance after you have launched it? (Select two)

ANS: **You can change the tenancy of an instance from dedicated to host**

**You can change the tenancy of an instance from host to dedicated**

A startup has created a new web application for users to complete a risk assessment survey for COVID-19 symptoms via a self-administered questionnaire. The startup has purchased the domain covid19survey.com using Route 53. The web development team would like to create a Route 53 record so that all traffic for covid19survey.com is routed to www.covid19survey.com.

As a solutions architect, which of the following is the MOST cost-effective solution that you would recommend to the web development team?

ANS: **Create an alias record for covid19survey.com that routes traffic to www.covid19survey.com**

Alias records provide a Route 53–specific extension to DNS functionality. Alias records let you route traffic to selected AWS resources, such as CloudFront distributions and Amazon S3 buckets.

You can create an alias record at the top node of a DNS namespace, also known as the zone apex, however, you cannot create a CNAME record for the top node of the DNS namespace. So, if you register the DNS name covid19survey.com, the zone apex is covid19survey.com. You can't create a CNAME record for covid19survey.com, but you can create an alias record for covid19survey.com that routes traffic to www.covid19survey.com.

Exam Alert:

You should also note that Route 53 doesn't charge for alias queries to AWS resources but Route 53 does charge for CNAME queries. **Additionally, an alias record can only redirect queries to selected AWS resources such as S3 buckets, CloudFront distributions, and another record in the same Route 53 hosted zone; however a CNAME record can redirect DNS queries to any DNS record.** So, you can create a CNAME record that redirects queries from app.covid19survey.com to app.covid19survey.net.

A big data analytics company is working on a real-time vehicle tracking solution. The data processing workflow involves both I/O intensive and throughput intensive database workloads. The development team needs to store this real-time data in a NoSQL database hosted on an EC2 instance and needs to support up to 25,000 IOPS per volume.

As a solutions architect, which of the following EBS volume types would you recommend for this use-case?

ANS: **Provisioned IOPS SSD (io1)**

Provisioned IOPS SSD (io1) is backed by solid-state drives (SSDs) and is a high-performance EBS storage option designed for critical, I/O intensive database and application workloads, as well as throughput-intensive database workloads. io1 is designed to deliver a consistent baseline performance of up to 50 IOPS/GB to a maximum of 64,000 IOPS and provide up to 1,000 MB/s of throughput per volume. Therefore, the io1 volume type would be able to meet the requirement of 25,000 IOPS per volume for the given use-case.

Incorrect options:

**General Purpose SSD (gp2)** - gp2 is backed by solid-state drives (SSDs) and is suitable for a broad range of transactional workloads, including dev/test environments, low-latency interactive applications, and boot volumes. It supports max IOPS/Volume of 16,000.

The engineering team at a social media company wants to use Amazon CloudWatch alarms to automatically recover EC2 instances if they become impaired. The team has hired you as a solutions architect to provide subject matter expertise.

As a solutions architect, which of the following statements would you identify as CORRECT regarding this automatic recovery process? (Select two)

ANS: **A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata**

**If your instance has a public IPv4 address, it retains the public IPv4 address after recovery**

You can create an Amazon CloudWatch alarm to automatically recover the Amazon EC2 instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. Terminated instances cannot be recovered. A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata. If the impaired instance is in a placement group, the recovered instance runs in the placement group. If your instance has a public IPv4 address, **it retains the public IPv4** address after recovery. During instance recovery, the instance is migrated during an instance reboot, and **any data that is in-memory is lost.**

A global manufacturing company with facilities in the US, Europe, and Asia is designing a new distributed application to optimize its procurement workflow. The orders booked on one continent should be visible to all AWS Regions in a second or less. The database should be able to facilitate failover with a short Recovery Time Objective (RTO). The uptime of the application is critical to ensure that the manufacturing processes are not impacted.

As a solutions architect, which of the following will you recommend as the MOST cost-effective solution?

ANS: **Provision Amazon Aurora Global Database**

**INCORRECT : Provision Amazon DynamoDB global tables**

DynamoDB global tables is a much costlier solution than Aurora Global Database for the given requirement.

An AWS Organization is using Service Control Policies (SCP) for central control over the maximum available permissions for all accounts in their organization. This allows the organization to ensure that all accounts stay within the organization’s access control guidelines.

Which of the given scenarios are correct regarding the permissions described below? (Select three)

ANS: **If a user or role has an IAM permission policy that grants access to an action that is either not allowed or explicitly denied by the applicable SCPs, the user or role can't perform that action**

**SCPs affect all users and roles in attached accounts, including the root user**

**SCPs do not affect service-linked role**

An e-commerce company runs its web application on EC2 instances in an Auto Scaling group and it's configured to handle consumer orders in an SQS queue for downstream processing. The DevOps team has observed that the performance of the application goes down in case of a sudden spike in orders received.

As a solutions architect, which of the following solutions would you recommend to address this use-case?

ANS:

**Use a target tracking scaling policy based on a custom Amazon SQS queue metric**

A retail company has a fleet of EC2 instances running behind an Auto Scaling group (ASG). The development team has configured two metrics that control the scale-in and scale-out policies of ASG. First one is a target tracking policy that uses a custom metric to add and remove two new instances, based on the number of SQS messages in the queue. The other is a step scaling policy that uses the CloudWatch CPU Utilization metric to launch one new instance when the existing instance exceeds 90 percent utilization for a specified length of time.

While testing, the scale-out policy criteria for both policies was met at the same time. How many new instances will be launched because of these multiple scaling policies?

ANS: **Amazon EC2 Auto Scaling chooses the policy that provides the largest capacity, so policy with the custom metric is triggered, and two new instances will be launched by the ASG**

When these situations occur, Amazon EC2 Auto Scaling chooses the policy that provides the largest capacity for both scale-out and scale-in. Suppose, for example, that the policy for CPUUtilization launches one instance, while the policy for the SQS queue launches two instances. If the scale-out criteria for both policies are met at the same time, Amazon EC2 Auto Scaling gives precedence to the SQS queue policy. This results in the Auto Scaling group launching two instances. The approach of giving precedence to the policy that provides the largest capacity applies even when the policies use different criteria for scaling in.

AWS recommends caution when using target tracking scaling policies with step scaling policies because conflicts between these policies can cause undesirable behavior. For example, if the step scaling policy initiates a scale-in activity before the target tracking policy is ready to scale in, the scale-in activity will not be blocked. After the scale-in activity completes, the target tracking policy could instruct the group to scale out again.

A global pharmaceutical company wants to move most of the on-premises data into Amazon S3, Amazon EFS, and Amazon FSx for Windows File Server easily, quickly, and cost-effectively.

As a solutions architect, which of the following solutions would you recommend as the BEST fit to automate and accelerate online data transfers to these AWS storage services?

ANS: **Use AWS DataSync to automate and accelerate online data transfers to the given AWS storage services**

AWS DataSync is an online data transfer service that simplifies, automates, and accelerates copying large amounts of data to and from AWS storage services over the internet or AWS Direct Connect.

AWS DataSync fully automates and accelerates moving large active datasets to AWS, up to 10 times faster than command-line tools. It is natively integrated with Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon CloudWatch, and AWS CloudTrail, which provides seamless and secure access to your storage services, as well as detailed monitoring of the transfer.

DataSync uses a purpose-built network protocol and scale-out architecture to transfer data. A single DataSync agent is capable of saturating a 10 Gbps network link.

DataSync fully automates the data transfer. It comes with retry and network resiliency mechanisms, network optimizations, built-in task scheduling, monitoring via the DataSync API and Console, and CloudWatch metrics, events, and logs that provide granular visibility into the transfer process. DataSync performs data integrity verification both during the transfer and at the end of the transfer.

A company uses Application Load Balancers (ALBs) in multiple AWS Regions. The ALBs receive inconsistent traffic that varies throughout the year. The engineering team at the company needs to allow the IP addresses of the ALBs in the on-premises firewall to enable connectivity.

Which of the following represents the MOST scalable solution with minimal configuration changes?

ANS: **Set up AWS Global Accelerator. Register the ALBs in different Regions to the Global Accelerator. Configure the on-premises firewall's rule to allow static IP addresses associated with the Global Accelerator** –

Associate the static IP addresses provided by AWS Global Accelerator to regional AWS resources or endpoints, such as Network Load Balancers, Application Load Balancers, EC2 Instances, and Elastic IP addresses. The IP addresses are anycast from AWS edge locations so they provide onboarding to the AWS global network close to your users.

Simplified and resilient traffic routing for multi-Region applications using Global Accelerator:

INCORRECT: **Set up a Network Load Balancer (NLB) in one Region. Register the private IP addresses of the ALBs in different Regions with the NLB. Configure the on-premises firewall's rule to allow the Elastic IP address attached to the NLB** - Using a single NLB is not possible across AWS regions since an NLB is Region bound. Multiple NLBs have to be registered for the on-premises firewall.

Your company has deployed an application that will perform a lot of overwrites and deletes on data and require the latest information to be available anytime data is read via queries on database tables.

As a Solutions Architect, which database technology will you recommend?

ANS: **Amazon Relational Database Service (Amazon RDS)** -All RDS transactions must be ACID compliant or be Atomic, Consistent, Isolated, and Durable to ensure data integrity.

Atomicity requires that either transaction as a whole is successfully executed or if a part of the transaction fails, then the entire transaction be invalidated.

Incorrect : s3; s3 is not a proper DB service with tables

---------------------------------------------------------------------------------------------------------------------------

A company wants an easy way to deploy and manage large fleets of Snowball devices. Which of the following solutions can be used to address the given requirements?

**ANS: AWS OpsHub**

AWS OpsHub is a graphical user interface you can use to manage your AWS Snowball devices, enabling you to rapidly deploy edge computing workloads and simplify data migration to the cloud. With just a few clicks in AWS OpsHub, you have the full functionality of the Snowball devices at your fingertips; you can unlock and configure devices, drag-and-drop data to devices, launch applications, and monitor device metrics. AWS OpsHub supports the Snowball Edge Storage Optimized and Snowball Edge Compute Optimized devices.

A company runs a popular dating website on the AWS Cloud. As a Solutions Architect, you've designed the architecture of the website to follow a serverless pattern on the AWS Cloud using API Gateway and AWS Lambda. The backend uses an RDS PostgreSQL database. Currently, the application uses a username and password combination to connect the Lambda function to the RDS database.

You would like to improve the security at the authentication level by leveraging short-lived credentials. What will you choose? (Select two)

ANS: **Use IAM authentication from Lambda to RDS PostgreSQL**

**Attach an AWS Identity and Access Management (IAM) role to AWS Lambda**

INCORRECT:

**Restrict the RDS database security group to the Lambda's security group**

**Deploy AWS Lambda in a VPC**

This question is very tricky because all answers do indeed increase security. But the question is related to authentication mechanisms, and as such, deploying a Lambda in a VPC or tightening security groups does not change the authentication layer.

A CRM company has a SaaS (Software as a Service) application that feeds updates to other in-house and third-party applications. The SaaS application and the in-house applications are being migrated to use AWS services for this inter-application communication.

As a Solutions Architect, which of the following would you suggest to asynchronously decouple the architecture?

ANS: **Use Amazon EventBridge to decouple the system architecture** –

Amazon EventBridge is recommended when you want to build an application that reacts to events from SaaS applications and/or AWS services. Amazon **EventBridge is the only event-based service that integrates directly with third-party SaaS partners.** Amazon EventBridge also automatically ingests events from over 90 AWS services without requiring developers to create any resources in their account. Further, Amazon EventBridge uses a defined JSON-based structure for events and allows you to create rules that are applied across the entire event body to select events to forward to a target. Amazon EventBridge currently supports over 15 AWS services as targets, including AWS Lambda, Amazon SQS, Amazon SNS, and Amazon Kinesis Streams and Firehose, among others. At launch, Amazon EventBridge is has limited throughput (see Service Limits) which can be increased upon request, and typical latency of around half a second.

As an e-sport tournament hosting company, you have servers that need to scale and be highly available. Therefore you have deployed an Elastic Load Balancer (ELB) with an Auto Scaling group (ASG) across 3 Availability Zones (AZs). When e-sport tournaments are running, the servers need to scale quickly. And when tournaments are done, the servers can be idle. As a general rule, you would like to be highly available, have the capacity to scale and optimize your costs.

What do you recommend? (Select two)

ANS: **Set the minimum capacity to 2**

even though our ASG is deployed across 3 AZs, the minimum capacity to be highly available is 2. When we specify 2 as the minimum capacity, the ASG would create these 2 instances in separate AZs. If demand goes up, the ASG would spin up a new instance in the third AZ. Later as the demand subsides, the ASG would scale-in and the instance count would be back to 2.

**Set the minimum capacity to 3** - This is not a cost-effective option, as two instances in two different AZs are enough to make the architecture disaster-proof.

Reporters at a news agency upload/download video files (about 500MB each) to/from an S3 bucket as part of their daily work. As the agency has started offices in remote locations, it has resulted in poor latency for uploading and accessing data to/from S3. The agency wants to continue using S3 but wants to improve the performance.

As a solutions architect, which of the following solutions do you propose to address this issue? (Select two)

ANS: **Use Amazon CloudFront distribution with origin as the S3 bucket. This would speed up uploads as well as downloads for the video files**

Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, high transfer speeds, within a developer-friendly environment. When an object from S3 that is set up with CloudFront CDN is requested, the request would come through the Edge Location transfer paths only for the first request. Thereafter, it would be served from the nearest edge location to the users until it expires. So in this way, you can speed up uploads as well as downloads for the video files.

A health-care company manages its web application on Amazon EC2 instances running behind Auto Scaling group (ASG). The company provides ambulances for critical patients and needs the application to be reliable. The workload of the company can be managed on 2 EC2 instances and can peak up to 6 instances when traffic increases.

As a Solutions Architect, which of the following configurations would you select as the best fit for these requirements?

ANS: **The ASG should be configured with the minimum capacity set to 4, with 2 instances each in two different Availability Zones. The maximum capacity of the ASG should be set to 6**

Amazon EC2 Auto Scaling attempts to distribute instances evenly between the Availability Zones that are enabled for your Auto Scaling group. This is why the minimum capacity should be 4 instances and not 2. ASG will launch 2 instances each in both the AZs and this redundancy is needed to keep the service available always.

A financial services company is moving its IT infrastructure to AWS Cloud and wants to enforce adequate data protection mechanisms on Amazon S3 to meet compliance guidelines. The engineering team has hired you as a solutions architect to build a solution for this requirement.

Can you help the team identify the INCORRECT option from the choices below?

ANS: **S3 can encrypt object metadata by using Server-Side Encryption**

Amazon S3 is a simple key-value store designed to store as many objects as you want. You store these objects in one or more buckets, and each object can be up to 5 TB in size.

An object consists of the following:

Key – The name that you assign to an object. You use the object key to retrieve the object.

Version ID – Within a bucket, a key and version ID uniquely identify an object.

Value – The content that you are storing.

Metadata – A set of name-value pairs with which you can store information regarding the object.

Subresources – Amazon S3 uses the subresource mechanism to store object-specific additional information.

Access Control Information – You can control access to the objects you store in Amazon S3.

Metadata, which can be included with the object, is not encrypted while being stored on Amazon S3. Therefore, AWS recommends that customers not place sensitive information in Amazon S3 metadata.

A mobile chat application uses DynamoDB as its database service to provide low latency chat updates. A new developer has joined the team and is reviewing the configuration settings for DynamoDB which have been tweaked for certain technical requirements. CloudTrail service has been enabled on all the resources used for the project. Yet, DynamoDB encryption details are nowhere to be found.

Which of the following options can explain the root cause for the given issue?

ANS: **By default, all DynamoDB tables are encrypted under an AWS owned customer master key (CMK), which do not write to CloudTrail logs** - AWS owned CMKs are a collection of CMKs that an AWS service owns and manages for use in multiple AWS accounts. Although AWS owned CMKs are not in your AWS account, an AWS service can use its AWS owned CMKs to protect the resources in your account.

You do not need to create or manage the AWS owned CMKs. However, you cannot view, use, track, or audit them. You are not charged a monthly fee or usage fee for AWS owned CMKs and they do not count against the AWS KMS quotas for your account.

The key rotation strategy for an AWS owned CMK is determined by the AWS service that creates and manages the CMK.

All DynamoDB tables are encrypted. There is no option to enable or disable encryption for new or existing tables. By default, all tables are encrypted under an AWS owned customer master key (CMK) in the DynamoDB service account. However, you can select an option to encrypt some or all of your tables under a customer-managed CMK or the AWS managed CMK for DynamoDB in your account.

The engineering team at a weather tracking company wants to enhance the performance of its relation database and is looking for a caching solution that supports geospatial data.

As a solutions architect, which of the following solutions will you suggest?

ANS: **Use Amazon ElastiCache for Redis**

Redis now delivers sub-millisecond response times enabling millions of requests per second for real-time applications in Gaming, Ad-Tech, Financial Services, Healthcare, and IoT. Redis is a popular choice for caching, session management, gaming, leaderboards, real-time analytics, geospatial, ride-hailing, chat/messaging, media streaming, and pub/sub apps.

 In comparison to traditional disk based databases where most operations require a roundtrip to disk, in-memory data stores such as Redis don’t suffer the same penalty. They can therefore support an order of magnitude more operations and faster response times. The result is – blazing fast performance with average read or write operations taking less than a millisecond and support for millions of operations per second.

Redis has purpose-built commands for working with real-time geospatial data at scale. You can perform operations like finding the distance between two elements (for example people or places) and finding all elements within a given distance of a point.

A streaming solutions company is building a video streaming product by using an Application Load Balancer (ALB) that routes the requests to the underlying EC2 instances. The engineering team has noticed a peculiar pattern. The ALB removes an instance from its pool of healthy instances whenever it is detected as unhealthy but the Auto Scaling group fails to kick-in and provision the replacement instance.

What could explain this anomaly?

ANS: **The Auto Scaling group is using EC2 based health check and the Application Load Balancer is using ALB based health check**

A company wants to publish an event into an SQS queue whenever a new object is uploaded on S3.

Which of the following statements are true regarding this functionality?

ANS: **Only Standard SQS queue is allowed as an Amazon S3 event notification destination, whereas FIFO SQS queue is not allowed**

A company hires experienced specialists to analyze the customer service calls attended by its call center representatives. Now, the company wants to move to AWS Cloud and is looking at an automated solution to analyze customer service calls for sentiment analysis and security.

As a Solutions Architect, which of the following solutions would you recommend?

ANS: **Use Amazon Transcribe to convert audio files to text and Amazon Athena to understand the underlying customer sentiments** - Amazon Transcribe is an automatic speech recognition (ASR) service that makes it easy to convert audio to text. One key feature of the service is called speaker identification, which you can use to label each individual speaker when transcribing multi-speaker audio files. You can specify Amazon Transcribe to identify 2–10 speakers in the audio clip.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run. To leverage Athena, you can simply point to your data in Amazon S3, define the schema, and start querying using standard SQL. Most results are delivered within seconds.

Incorrect options:

**Use Kinesis Data Streams to read the audio files and machine learning (ML) algorithms to convert the audio files into text and run customer sentiment analysis** - Amazon Kinesis can be used to stream real-time data for further analysis and storage. Kinesis Data Streams cannot read audio files. You will still need to use AWS Transcribe for ASR services.

**Use Kinesis Data Streams to read the audio files and Amazon Alexa to convert them into text. Kinesis Data Analytics can be used to analyze these files and Amazon Quicksight can be used to visualize and display the output** - Kinesis Data Streams cannot read audio files. Amazon Alexa cannot be used as an Automatic Speech Recognition (ASR) service, though Alexa internally uses ASR for its working.

**Use Amazon Transcribe to convert audio files to text and Amazon Quicksight to run analysis on these text files to understand the underlying patterns. Visualize and display them onto user Dashboards for human analysis** - Amazon Quicksight is for the visual representation of data through Dashboards, graphs and various other modes. It has a rich feature set that helps analyze data and the complex relationships that exist between different data features. It is, however, not a powerful analysis tool like Amazon Athena.

A junior developer is learning to build websites using HTML, CSS, and JavaScript. He has created a static website and then deployed it on Amazon S3. Now he can't seem to figure out the endpoint for his super cool website. As a solutions architect, can you help him figure out the allowed formats for the Amazon S3 website endpoints? (Select two)

ANS: <http://bucket-name.s3-website.Region.amazonaws.com>

<http://bucket-name.s3-website-Region.amazonaws.com>: Depending on your Region, your Amazon S3 website endpoints follow one of these two formats.

s3-website dash (-) Region ‐ http://bucket-name.s3-website.Region.amazonaws.com

s3-website dot (.) Region ‐ http://bucket-name.s3-website-Region.amazonaws.com

These URLs return the default index document that you configure for the website.

A development team is looking for a solution that saves development time and deployment costs for an application that uses a high-throughput request-response message pattern.

Which of the following SQS queue types is the best fit to meet this requirement?

ANS: **Amazon SQS temporary queues** - Temporary queues help you save development time and deployment costs when using common message patterns such as request-response. You can use the Temporary Queue Client to create high-throughput, cost-effective, application-managed temporary queues.

The client maps multiple temporary queues—application-managed queues created on demand for a particular process—onto a single Amazon SQS queue automatically. This allows your application to make fewer API calls and have a higher throughput when the traffic to each temporary queue is low. When a temporary queue is no longer in use, the client cleans up the temporary queue automatically, even if some processes that use the client aren't shut down cleanly.

The following are the benefits of temporary queues:

1. They serve as lightweight communication channels for specific threads or processes.
2. They can be created and deleted without incurring additional costs.
3. They are API-compatible with static (normal) Amazon SQS queues. This means that existing code that sends and receives messages can send messages to and receive messages from virtual queues.

To better support short-lived, lightweight messaging destinations, AWS recommends Amazon SQS Temporary Queue Client. This client makes it easy to create and delete many temporary messaging destinations without inflating your AWS bill. The key concept behind the client is the virtual queue. Virtual queues let you multiplex many low-traffic queues onto a single SQS queue. Creating a virtual queue only instantiates a local buffer to hold messages for consumers as they arrive; there is no API call to SQS and no costs associated with creating a virtual queue.

The engineering team at a multi-national company uses AWS Firewall Manager to centrally configure and manage firewall rules across its accounts and applications using AWS Organizations.

Which of the following AWS resources can the AWS Firewall Manager configure rules on? (Select three)

ANS: **AWS WAF**

**AWS Shield Advanced**

**VPC Security Groups**

AWS Firewall Manager is a security management service which allows you to centrally configure and manage firewall rules across your accounts and applications in AWS Organizations. As new applications are created, Firewall Manager makes it easy to bring new applications and resources into compliance by enforcing a common set of security rules. Now you have a single service to build firewall rules, create security policies, and enforce them in a consistent, hierarchical manner across your entire infrastructure.

Using AWS Firewall Manager, you can centrally configure AWS WAF rules, AWS Shield Advanced protection, Amazon Virtual Private Cloud (VPC) security groups, AWS Network Firewalls, and Amazon Route 53 Resolver DNS Firewall rules across accounts and resources in your organization. It does not support Network ACLs as of today.

An engineering team wants to orchestrate multiple Amazon ECS task types running on EC2 instances that are part of the ECS cluster. The output and state data for all tasks need to be stored. The amount of data output by each task is approximately 20 MB and there could be hundreds of tasks running at a time. As old outputs are archived, the storage size is not expected to exceed 1 TB.

As a solutions architect, which of the following would you recommend as an optimized solution for high-frequency reading and writing?

ANS: **Use Amazon EFS with Provisioned Throughput mode** - Provisioned Throughput mode is available for applications with high throughput to storage (MiB/s per TiB) ratios, or with requirements greater than those allowed by the Bursting Throughput mode. For example, say you're using Amazon EFS for development tools, web serving, or content management applications where the amount of data in your file system is low relative to throughput demands. Your file system can now get the high levels of throughput your applications require without having to pad your file system.

INCORRECT:  
**Use Amazon EFS with Bursting Throughput mode** - With Bursting Throughput mode, a file system's throughput scales as the amount of data stored in the standard storage class grows. File-based workloads are typically spiky, driving high levels of throughput for short periods of time, and low levels of throughput the rest of the time. To accommodate this, Amazon EFS is designed to burst to high throughput levels for periods of time.

The use-case mentions that the solution should be optimized for high-frequency reading and writing even when the old outputs are archived, therefore Provisioned Throughput mode is a better fit as it guarantees high levels of throughput your applications require without having to pad your file system

**Use an Amazon EBS volume mounted to the ECS cluster instances** - EFS has a higher throughput than EBS.

During a review, a security team has flagged concerns over an Amazon EC2 instance querying IP addresses used for cryptocurrency mining. The EC2 instance does not host any authorized application related to cryptocurrency mining.

Which AWS service can be used to protect the EC2 instances from such unauthorized behavior in the future?

ANS: **Amazon GuardDuty** - Amazon GuardDuty continuously monitors for malicious or unauthorized behavior to help protect your AWS resources, including your AWS accounts and access keys. GuardDuty identifies any unusual or unauthorized activity, like cryptocurrency mining or infrastructure deployments in a region that has never been used. Powered by threat intelligence and machine learning, GuardDuty is continuously evolving to help you protect your AWS environment.

The cryptocurrency finding expands the service’s ability to detect Amazon EC2 instances querying IP addresses associated with the cryptocurrency-related activity. The finding type is: CryptoCurrency:EC2/BitcoinTool.B, CryptoCurrency:EC2/BitcoinTool.B!DNS.

This finding informs you that the listed EC2 instance in your AWS environment is querying a domain name that is associated with Bitcoin or other cryptocurrency-related activity. Bitcoin is a worldwide cryptocurrency and digital payment system that can be exchanged for other currencies, products, and services. Bitcoin is a reward for bitcoin mining and is highly sought after by threat actors.

If you use the EC2 instance to mine or manage cryptocurrency, or this instance is otherwise involved in blockchain activity, this finding could represent expected activity for your environment. If this is the case in your AWS environment, AWS recommends that you set up a suppression rule for this finding.

A company has noticed several provisioned throughput exceptions on its DynamoDB database due to major spikes in the writes to the database. The development team wants to decouple the application layer from the database layer and dedicate a worker process to writing the data to DynamoDB.

Which middleware do you recommend on using that can scale infinitely and meet these requirements?

ANS: **Amazon Simple Queue Service (SQS)**

**INCORRECT:**

**Kinesis Data Streams**

Kinesis cannot scale infinitely and we may have the same throughput issues. Therefore this option is not the right fit for the given use-case.

**Amazon Simple Notification Service (SNS)** -SNS won't keep our data if it cannot be delivered, so this option is incorrect.

A company is deploying a web application and it wants to ensure that only the web tier of the application is publicly accessible. To accomplish this, the engineering team has designed the VPC with a public subnet and a private subnet. The application will be hosted on several EC2 instances in an Auto Scaling group. The team also wants TLS termination to be offloaded from the EC2 instances.

Which solution should a solutions architect implement to address these requirements?

**ANS: Set up a Network Load Balancer in the public subnet. Create an Auto Scaling group in the private subnet and associate it with the Network Load Balancer**

With Network Load Balancer (NLB), you can offload the decryption/encryption of TLS traffic from your application servers to the Network Load Balancer, which helps you optimize the performance of your backend application servers while keeping your workloads secure. Additionally, Network Load Balancers preserve the source IP of the clients to the back-end applications, while terminating TLS on the load balancer.

Exam Alert:

You should note that the Application Load Balancer also supports TLS offloading. The Classic Load Balancer supports SSL offloading.

The engineering team at a company wants to create a daily big data analysis job leveraging Spark for analyzing online/offline sales and customer loyalty data to create customized reports on a client-by-client basis. The big data analysis job needs to read the data from Amazon S3 and output it back to S3.

Which technology do you recommend to run the Big Data analysis job? (Select two)

ANS: **Amazon EMR** - Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. With EMR you can run Petabyte-scale analysis at less than half of the cost of traditional on-premises solutions and over 3x faster than standard Apache Spark. EMR is used for launching Hadoop / Spark clusters. For short-running jobs, you can spin up and spin down clusters and pay per second for the instances used. For long-running workloads, you can create highly available clusters that automatically scale to meet demand. Amazon EMR uses Hadoop, an open-source framework, to distribute your data and processing across a resizable cluster of Amazon EC2 instances.

**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 ETL jobs can use Amazon S3, data stores in a VPC, or on-premises JDBC data stores as a source. AWS Glue jobs extract data, transform it, and load the resulting data back to S3, data stores in a VPC, or on-premises JDBC data stores as a target.

The systems administrator at a company wants to set up a highly available architecture for a bastion host solution.

As a solutions architect, which of the following options would you recommend as the solution?

ANS: Network Load Balancer is best suited for use-cases involving low latency and high throughput workloads that involve scaling to millions of requests per second. Network Load Balancer operates at the connection level (Layer 4), routing connections to targets - Amazon EC2 instances, microservices, and containers – within Amazon Virtual Private Cloud (Amazon VPC) based on IP protocol data.

INCORRECT: **Create a public Application Load Balancer that links to EC2 instances that are bastion hosts managed by an ASG** -An ALB only supports HTTP traffic, which is layer 7, while the SSH protocol is based on TCP and is layer 4. So, the Application Load Balancer doesn't work.

A Big Data consulting company runs large distributed and replicated workloads on the on-premises data center. The company now wants to move these workloads to Amazon EC2 instances by using the placement groups feature and it wants to minimize correlated hardware failures.

Which of the following represents the correct placement group configuration for the given requirement?

ANS: **Partition placement groups** - Partition placement groups help reduce the likelihood of correlated hardware failures for your application.

Partition placement groups can be used to deploy large distributed and replicated workloads, such as HDFS, HBase, and Cassandra, across distinct racks. When you launch instances into a partition placement group, Amazon EC2 tries to distribute the instances evenly across the number of partitions that you specify. You can also launch instances into a specific partition to have more control over where the instances are placed.

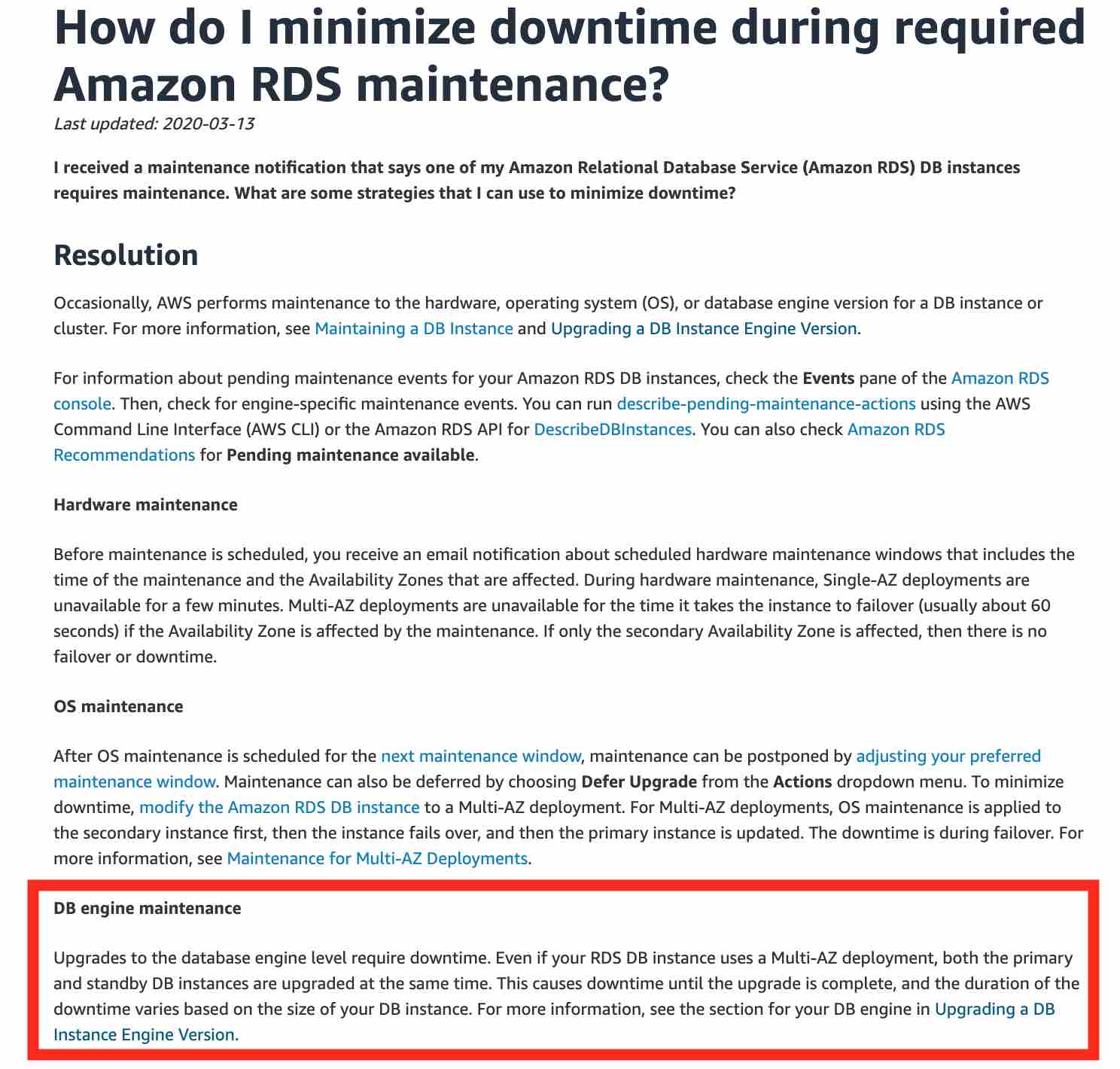
A partition placement group can have partitions in multiple Availability Zones in the same Region. A partition placement group can have a maximum of seven partitions per Availability Zone

INCORRECT: Spread placement groups: As the use-case talks about running large distributed and replicated workloads, so it needs more instances, therefore this option is not the right fit for the given use-case.

The DevOps team at a major financial services company uses Multi-Availability Zone (Multi-AZ) deployment for its MySQL RDS database in order to automate its database replication and augment data durability. The DevOps team has scheduled a maintenance window for a database engine level upgrade for the coming weekend.

Which of the following is the correct outcome during the maintenance window?

ANS: **Any database engine level upgrade for an RDS DB instance with Multi-AZ deployment triggers both the primary and standby DB instances to be upgraded at the same time. This causes downtime until the upgrade is complete**



Your company is building a video streaming service accessible to users who have paid an ongoing subscription. The subscription data is stored in DynamoDB. You would like to expose the users to a serverless architecture allowing them to request the video files that sit on Amazon S3 and are distributed by CloudFront and protected by an origin access identity (OAI).

Which of the following options can be combined to build a solution? (Select two)

ANS: **Use AWS Lambda to generate the URL**

**Generate a CloudFront signed URL**

Create a special CloudFront user called an origin access identity (OAI) and associate it with your distribution. Configure your S3 bucket permissions so that CloudFront can use the OAI to access the files in your bucket and serve them to your users. Make sure that users can’t use a direct URL to the S3 bucket to access a file there.

After you take these steps, users can only access your files through CloudFront, not directly from the S3 bucket. In general, if you’re using an Amazon S3 bucket as the origin for a CloudFront distribution, you can either allow everyone to have access to the files there, or you can restrict access. If you restrict access by using, for example, CloudFront signed URLs or signed cookies, you also won’t want people to be able to view files by simply using the direct Amazon S3 URL for the file. Instead, you want them to only access the files by using the CloudFront URL, so your content remains protected.

To generate this URL we must code, and Lambda is the perfect tool for running that code on the fly.

INCORRECT: **Generate an S3 pre-signed URL** - Generating S3 pre-signed URLs would bypass CloudFront, therefore we should use CloudFront signed URLs.

The DevOps team at an e-commerce company has deployed a fleet of EC2 instances under an Auto Scaling group (ASG). The instances under the ASG span two Availability Zones (AZ) within the us-east-1 region. All the incoming requests are handled by an Application Load Balancer (ALB) that routes the requests to the EC2 instances under the ASG. As part of a test run, two instances (instance 1 and 2, belonging to AZ A) were manually terminated by the DevOps team causing the Availability Zones to become unbalanced. Later that day, another instance (belonging to AZ B) was detected as unhealthy by the Application Load Balancer's health check.

Can you identify the correct outcomes for these events? (Select two)

ANS: **As the Availability Zones got unbalanced, Amazon EC2 Auto Scaling will compensate by rebalancing the Availability Zones. When rebalancing, Amazon EC2 Auto Scaling launches new instances before terminating the old ones, so that rebalancing does not compromise the performance or availability of your application**

**INCORRECT: As the Availability Zones got unbalanced, Amazon EC2 Auto Scaling will compensate by rebalancing the Availability Zones. When rebalancing, Amazon EC2 Auto Scaling terminates old instances before launching new instances, so that rebalancing does not cause extra instances to be launched**

The engineering team at an e-commerce company wants to set up a custom domain for internal usage such as internaldomainexample.com. The team wants to use the private hosted zones feature of Route 53 to accomplish this.

Which of the following settings of the VPC need to be enabled? (Select two)

ANS: **enableDnsHostnames**

**enableDnsSupport**

A development team has noticed that one of the EC2 instances has been incorrectly configured with the 'DeleteOnTermination' attribute set to True for its root EBS volume.

As a Solution's Architect, can you suggest a way to disable this flag while the instance is still running?

ANS: **Set the DeleteOnTermination attribute to False using the command line** - If the instance is already running, you can set DeleteOnTermination to False using the command line.

Incorrect options:

**Update the attribute using AWS management console. Select the EC2 instance and then uncheck the Delete On Termination check box for the root EBS volume** - You can set the DeleteOnTermination attribute to False when you launch a new instance. It is not possible to update this attribute of a running instance from the AWS console.

**Set the DisableApiTermination attribute of the instance using the API** - By default, you can terminate your instance using the Amazon EC2 console, command-line interface, or API. To prevent your instance from being accidentally terminated using Amazon EC2, you can enable termination protection for the instance. The DisableApiTermination attribute controls whether the instance can be terminated using the console, CLI, or API. This option cannot be used to control the delete status for the EBS volume when the instance terminates.

The engineering team at a social media company has noticed that while some of the images stored in S3 are frequently accessed, others sit idle for a considerable span of time.

As a solutions architect, what is your recommendation to build the MOST cost-effective solution?

ANS: **Store the images using the S3 Intelligent-Tiering storage class**

**INCORRECT:**

**Create a data monitoring application on an EC2 instance in the same region as the bucket storing the images. The application is triggered daily via CloudWatch and it changes the storage class of infrequently accessed objects to S3 Standard-IA and the frequently accessed objects are migrated to S3 Standard class**

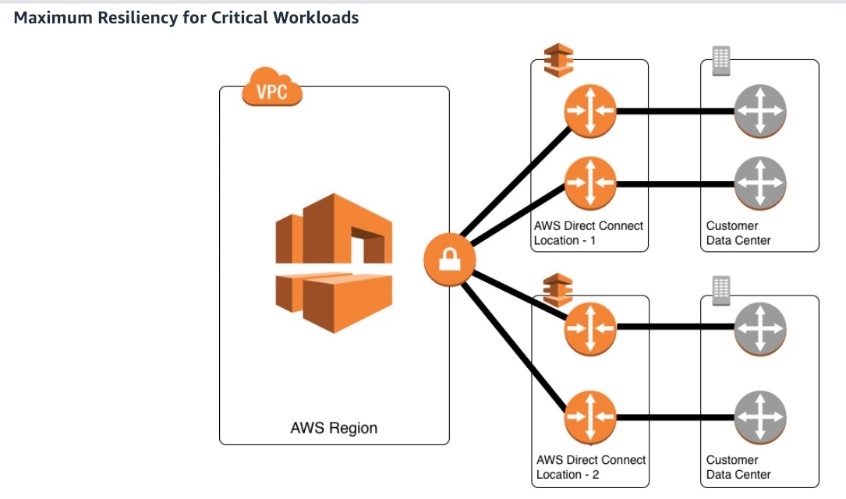
Creating a data monitoring application on an EC2 instance for managing the desired S3 storage class entails significant development cost as well as infrastructure maintenance effort. The S3 Intelligent-Tiering storage class does the job in a cost-effective way. Therefore both these options are incorrect.

To support critical production workloads that require maximum resiliency, a company wants to configure network connections between its Amazon VPC and the on-premises infrastructure. The company needs AWS Direct Connect connections with speeds greater than 1 Gbps.

As a solutions architect, which of the following will you suggest as the best architecture for this requirement?

ANS: **Opt for two separate Direct Connect connections terminating on separate devices in more than one Direct Connect location** - Maximum resilience is achieved by separate connections terminating on separate devices in more than one location. This configuration offers customers maximum resilience to failure. As shown in the figure above, such a topology provides resilience to device failure, connectivity failure, and complete location failure. You can use Direct Connect Gateway to access any AWS Region (except AWS Regions in China) from any AWS Direct Connect locations.

Maximum Resiliency for Critical Workloads:



A company runs an analytics workload with heavy reads and writes through the workload lifecycle. The development team at the company is interested in using Amazon S3 as the data lake to support this workload. The team has hired you as a solutions architect to advise them on the S3 data consistency model.

Which of the following statements would you identify as correct?

ANS: **Amazon S3 is strongly consistent for all GET, PUT and LIST operations**

INCORRECT: **Amazon S3 is strongly consistent for all GET, PUT and LIST operations and eventually consistent for operations that need metadata information** - For all existing and new objects, and in all regions, all S3 GET, PUT, and LIST operations, as well as operations that change object tags, ACLs, or metadata, are now strongly consistent. So this option is not correct.

You are deploying a critical monolith application that must be deployed on a single web server, as it hasn't been created to work in distributed mode. Still, you want to make sure your setup can automatically recover from the failure of an AZ

Which of the following options should be combined to form the MOST cost-efficient solution? (Select three)

ANS: **Create an auto-scaling group that spans across 2 AZ, which min=1, max=1, desired=1.** So we have an ASG with desired=1, across two AZ, so that if an instance goes down, it is automatically recreated in another AZ. So this option is correct.

**Create an Elastic IP and use the EC2 user-data script to attach it**

Application Load Balancer (ALB) operates at the request level (layer 7), routing traffic to targets – EC2 instances, containers, IP addresses, and Lambda functions based on the content of the request.

An Elastic IP address is a static IPv4 address designed for dynamic cloud computing. An Elastic IP address is associated with your AWS account. With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.

Now, between the ALB and the Elastic IP. If we use an ALB, things will still work, but we will have to pay for the provisioned ALB which sends traffic to only one EC2 instance. Instead, to minimize costs, we must use an Elastic IP

**Assign an EC2 Instance Role to perform the necessary API calls**

For that Elastic IP to be attached to our EC2 instance, we must use an EC2 user data script, and our EC2 instance must have the correct IAM permissions to perform the API call, so we need an EC2 instance role.

You have deployed a database technology that has a synchronous replication mode to survive disasters in data centers. The database is therefore deployed on two EC2 instances in two Availability Zones. The database must be publicly available so you have deployed the EC2 instances in public subnets. The replication protocol currently uses the EC2 public IP addresses.

What can you do to decrease the replication cost?

ANS: **Use the EC2 instances private IP for the replication**

The source of the cost is that traffic between two EC2 instances is going over the public internet, thus incurring high costs. Here, the correct answer is to use Private IP, so that the network remains private, for a minimal cost.

INCORRECT: **Create a Private Link between the two EC2 instances** - Private Link is a distractor in this question. Private Link is leveraged to create a private connection between 2 VPCs for an application that is fronted by an NLB in an account, and an Elastic Network Interface (ENI) in another account, without the need of VPC peering and allowing the connections between the two to remain within the AWS network.

Your application is deployed on EC2 instances fronted by an Application Load Balancer. Recently, your infrastructure has come under attack. Attackers perform over 100 requests per second, while your normal users only make about 5 requests per second.

How can you efficiently prevent attackers from overwhelming your application?

**ANS: Use a Web Application Firewall and setup a rate-based rule**

AWS WAF is a web application firewall that helps protect your web applications or APIs against common web exploits that may affect availability, compromise security, or consume excessive resources. AWS WAF gives you control over how traffic reaches your applications by enabling you to create security rules that block common attack patterns, such as SQL injection or cross-site scripting, and rules that filter out specific traffic patterns you define.

The correct answer is to use WAF (which has integration on top of your ALB) and define a rate-based rule.

**INCORRECT: Use AWS Shield Advanced and setup a rate-based rule:** AWS Shield Advanced will give you DDoS protection overall, and you cannot set up rate-based rules in Shield.

A social media application lets users upload photos and perform image editing operations. The application offers two classes of service: pro and lite. The product team wants the photos submitted by pro users to be processed before those submitted by lite users. Photos are uploaded to S3 and the job information is sent to Amazon SQS.

As a solutions architect, which of the following solutions would you recommend?

ANS: **Create two SQS standard queues: one for pro and one for lite. Set up EC2 instances to prioritize polling for the pro queue over the lite queue**

AWS recommends using separate queues to provide prioritization of work. Therefore, for the given use case, you need to create an SQS standard queue for processing pro users' photos and another SQS standard queue for processing lite users' photos. Then you can configure EC2 instances to prioritize polling for the pro queue over the lite queue.

Incorrect options:

**Create two SQS standard queues: one for pro and one for lite. Set the lite queue to use short polling and the pro queue to use long polling**

**Create two SQS FIFO queues: one for pro and one for lite. Set the lite queue to use short polling and the pro queue to use long polling**

Amazon SQS long polling is a way to retrieve messages from your Amazon SQS queues. While the regular short polling returns immediately, even if the message queue being polled is empty, long-polling doesn’t return a response until a message arrives in the message queue, or the long poll times out. Since long polling or short polling cannot impact the priority of processing for the two queues, so both these options are incorrect.

**Create one SQS standard queue. Set the visibility timeout of the pro photos to zero. Set up EC2 instances to prioritize visibility settings so pro photos are processed first** - To prevent other consumers from processing the message again, Amazon SQS sets a visibility timeout, a period of time during which Amazon SQS prevents other consumers from receiving and processing the message. The default visibility timeout for a message is 30 seconds. The minimum is 0 seconds. The maximum is 12 hours. Setting visibility timeout to zero can result in the same pro photo being processed by more than one consumer. This does not help in prioritizing the processing of pro photos over the lite photos.

A Big Data analytics company is using a fleet of Amazon EC2 instances to ingest Internet-of-Things (IoT) data from various data sources. The data is in JSON format and ingestion rates can be as high as 1 MB/s. When an EC2 instance is restarted, the in-flight data is lost. The analytics team at the company wants to store as well as query the ingested data in near-real-time.

Which of the following solutions provides near-real-time data querying that is scalable with minimal data loss?

ANS: **Capture data in Amazon Kinesis Data Firehose with Amazon Redshift as the destination. Use Amazon Redshift to query the data** –

Amazon Kinesis Data Firehose is the easiest way to capture, transform, and load streaming data into Redshift for near real-time analytics. It is also an auto-scaling solution as there is no need to provision any shards like Kinesis Data Streams.

Redshift allows you to run complex analytic queries against petabytes of structured data, using sophisticated query optimization, columnar storage on high-performance local disks, and massively parallel query execution. Most results come back in seconds.

INCORRECT: **Capture data in Amazon Kinesis Data Streams. Use Kinesis Data Analytics to query and analyze this streaming data in real-time** - For Kinesis Data Streams, you have to manually allocate the shards for scaling the data ingestion process. Kinesis Data Streams (KDS) and Kinesis Data Analytics are for real-time processing of data and cannot provide long-term storage of data unlike a database or a data warehouse. So, this option is not right for the current use case.

A photo-sharing company is storing user profile pictures in an S3 bucket and an image analysis application is deployed on four EC2 instances. A solutions architect would like to trigger an image analysis procedure only on one of the four EC2 instances for each photo uploaded.

What do you recommend?

ANS: **Create an S3 Event Notification that sends a message to an SQS queue. Make the EC2 instances read from the SQS queue**

Amazon S3 supports the following destinations where it can publish events:

Amazon Simple Notification Service (Amazon SNS) topic

Amazon Simple Queue Service (Amazon SQS) queue

AWS Lambda

By using SQS, we know only one EC2 instance among the four will pick up a message and process it.

INCORRECT:

**Subscribe the EC2 instances to the S3 Inventory stream** - S3 Inventory is a distractor. If you're curious - Amazon S3 inventory helps you manage your storage by creating lists of the objects in an S3 bucket on a defined schedule.

**Create a CloudWatch Event that reacts to objects uploads in S3 and invokes one of the EC2 instances** - CloudWatch Events cannot invoke applications on EC2 instances, so we have to rule out that answer.

The engineering team at a startup is evaluating the most optimal block storage volume type for the EC2 instances hosting its flagship application. The storage volume should support very low latency but it does not need to persist the data when the instance terminates. As a solutions architect, you have proposed using Instance Store volumes to meet these requirements.

Which of the following would you identify as the key characteristics of the Instance Store volumes? (Select two)

ANS: **You can't detach an instance store volume from one instance and attach it to a different instance** - You can specify instance store volumes for an instance only when you launch it. You can't detach an instance store volume from one instance and attach it to a different instance. The data in an instance store persists only during the lifetime of its associated instance. If an instance reboots (intentionally or unintentionally), data in the instance store persists.

**If you create an AMI from an instance, the data on its instance store volumes isn't preserved** - If you create an AMI from an instance, the data on its instance store volumes isn't preserved and isn't present on the instance store volumes of the instances that you launch from the AMI.

Incorrect options:

**Instance store is reset when you stop or terminate an instance. Instance store data is preserved during hibernation** - When you stop, hibernate, or terminate an instance, every block of storage in the instance store is reset. Therefore, this option is incorrect.

**You can specify instance store volumes for an instance when you launch or restart it** - You can specify instance store volumes for an instance only when you launch it.

**An instance store is a network storage type** - An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer

An e-commerce website is migrating towards a microservices-based approach for their website and plans to expose their website from the same load balancer, linked to different target groups with different URLs: checkout.mycorp.com, www.mycorp.com, mycorp.com/products, and mycorp.com/orders. The website would like to use ECS on the backend to manage these microservices and possibly host the same container of the application multiple times on the same EC2 instance.

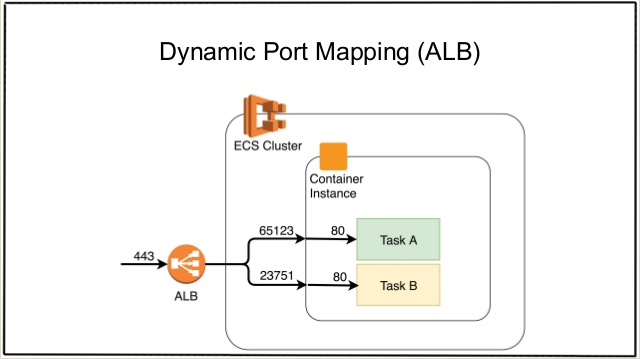
Which feature can help you achieve this with minimal effort?

ANS: **Application Load Balancer + dynamic port mapping**

Application Load Balancer can automatically distribute incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones.

Dynamic port mapping with an Application Load Balancer makes it easier to run multiple tasks on the same Amazon ECS service on an Amazon ECS cluster.

Dynamic port mapping **allows you to run multiple tasks over the same host using multiple random host ports** (in spite of defined host port). Dynamic port mapping with an Application Load Balancer makes it easier to run multiple tasks on the same Amazon ECS service on an Amazon ECS cluster



Which of the following S3 storage classes supports encryption by default for both data at rest as well as in-transit?

ANS: GLACIER

Incorrect options:

**Amazon S3 Standard-Infrequent Access/** **Amazon S3 Intelligent Tiering** / **Amazon S3 One Zone-IA:** It does not support encryption by default for both data at rest as well as in-transit.

A company uses a legacy on-premises reporting application that operates on gigabytes of .json files and represents years of data. The legacy application cannot handle the growing size of .json files. New .json files are added daily from various data sources to a central on-premises storage location. The company wants to continue to support the legacy application. The company has hired you as a solutions architect to build a solution that can manage ongoing data updates from your on-premises application to Amazon S3.

Which of the following solutions would you suggest to address the given requirement?

ANS: **Set up an on-premises file gateway. Configure data sources to write the .json files to the file gateway. Point the legacy analytics application to the file gateway. The file gateway should replicate the .json files to Amazon S3**

The file gateway employs a local read/write cache to provide low-latency access to data for file share clients in the same local area network (LAN) as the file gateway.

A company is developing a document management application on AWS. The application runs on EC2 instances in multiple Availability Zones. The company requires the document store to be highly available and the documents need to be returned immediately when requested. The engineering team has configured the application to use EBS to store the documents but the team is willing to consider other options to meet the availability requirement.

As a solutions architect, which of the following will you recommend?

ANS: **Set up Amazon EBS as the EC2 instance root volume and then configure the application to use S3 as the document store**

**Provision at least three Provisioned IOPS EBS volumes for the EC2 instances and then mount these volumes to the EC2 instances in a RAID 5 configuration** - RAID 5 and RAID 6 are not recommended for Amazon EBS because the parity write operations of these RAID modes consume some of the IOPS available to your volumes. Depending on the configuration of your RAID array, these RAID modes provide 20-30% fewer usable IOPS than a RAID 0 configuration. So this option is incorrect.

RAID configuration options

Creating a RAID 0 array allows you to achieve a higher level of performance for a file system than you can provision on a single Amazon EBS volume. Use RAID 0 when I/O performance is of the utmost importance. With RAID 0, I/O is distributed across the volumes in a stripe. If you add a volume, you get the straight addition of throughput and IOPS. However, keep in mind that performance of the stripe is limited to the worst performing volume in the set, and that the loss of a single volume in the set results in a complete data loss for the array.

The resulting size of a RAID 0 array is the sum of the sizes of the volumes within it, and the bandwidth is the sum of the available bandwidth of the volumes within it. For example, two 500 GiB io1 volumes with 4,000 provisioned IOPS each create a 1000 GiB RAID 0 array with an available bandwidth of 8,000 IOPS and 1,000 MiB/s of throughput.

Important

RAID 5 and RAID 6 are not recommended for Amazon EBS because the parity write operations of these RAID modes consume some of the IOPS available to your volumes. Depending on the configuration of your RAID array, these RAID modes provide 20-30% fewer usable IOPS than a RAID 0 configuration. Increased cost is a factor with these RAID modes as well; when using identical volume sizes and speeds, a 2-volume RAID 0 array can outperform a 4-volume RAID 6 array that costs twice as much.

RAID 1 is also not recommended for use with Amazon EBS. RAID 1 requires more Amazon EC2 to Amazon EBS bandwidth than non-RAID configurations because the data is written to multiple volumes simultaneously. In addition, RAID 1 does not provide any write performance improvement.

