Question – wise notes: AWS Mock – 2:

Question 1:

A company runs a data processing workflow that takes about 60 minutes to complete. The workflow can withstand disruptions and it can be started and stopped multiple times.

Correct Option:

Use Amazon EC2 Spot instances to run the workflow processes.

Explanation and Context:

Amazon EC2 Spot instances allow you to request spare Amazon EC2 computing capacity for up to 90% off the On-Demand price, making them ideal for processes that have flexible start and end times and can be interrupted. This matches the company's need as their workflow can handle disruptions and does not need to run continuously, thereby significantly reducing costs.

Incorrect Options:

1. **Use Amazon EC2 Reserved Instances**: While reserved instances provide a discount on the hourly charge for an instance, they are best suited for use cases where a consistent workload is expected. This option does not provide the flexibility of handling interruptions as cost-effectively as Spot instances.
2. **Use Amazon EC2 On-Demand Instances**: This option involves paying for compute capacity by the hour without long-term commitments, which would be more expensive than Spot instances for interruptible processes.
3. **Use AWS Lambda to run the workflow processes**: Lambda is designed for handling code execution in response to events and automatically managing the computing resources. However, for a 60-minute continuous workflow, managing the session state and execution context would be complex and potentially more costly compared to using EC2 Spot instances.

Conclusion / Important Points to Memorize:

- EC2 Spot instances provide significant cost savings for flexible and interruptible workflows.
- Reserved and On-Demand instances are less cost-effective for interruptible tasks.
  - AWS Lambda may introduce complexities and costs for long-running tasks.

Question 2:

A social photo-sharing company uses Amazon Simple Storage Service (Amazon S3) to store images uploaded by users. These images are encrypted in Amazon S3 using AWS Key Management Service (AWS KMS), and the company manages its own AWS KMS keys. A member of the DevOps team accidentally deleted the AWS KMS key, rendering the user's photo data unrecoverable. You are consulted to solve this crisis.

Correct Option:

As the AWS KMS key was deleted a day ago, it must be in the 'pending deletion' status, and you can just cancel the KMS key deletion and recover the key.

Explanation and Context:

AWS KMS allows you to delete keys by first marking them for deletion, during which you can set a waiting period (minimum 7 days, up to 30 days). If a key is deleted within this period, the deletion process can be canceled, allowing the key to be recovered. This fits the scenario as the key was deleted only a day ago.

Incorrect Options:

1. **Contact AWS Support to Retrieve the Key**: AWS support cannot recover a deleted KMS key if the deletion process has already been completed.
2. **The AWS KMS key can be recovered by the AWS root account user**: The root account user does not have the capability to recover a key once it is scheduled for deletion beyond the recovery period.
3. **The company should issue a notification on its web application about the loss of data**: This step would be unnecessary if the key can still be recovered.

Conclusion / Important Points to Memorize:

- Understand the deletion process for AWS KMS keys, including the ability to cancel pending deletions.
- Recognize the limitations of AWS support and root account capabilities in recovering deleted KMS keys.

Question 3:

A software engineering intern at an e-commerce company is documenting the process flow to provision Amazon EC2 instances via the Amazon EC2 API. These instances are to be used for an internal application that processes Human Resources payroll data. He wants to highlight those volume types that cannot be used as a boot volume. Can you help the intern by identifying those storage volume types that CANNOT be used as boot volumes while creating the instances? (Select two)

Correct Options:

1. **Throughput Optimized HDD (st1)**
2. **Cold HDD (sc1)**

Explanation and Context:

Both Throughput Optimized HDD and Cold HDD are designed for workloads that require large amounts of throughput at a lower cost. They are not suitable for boot volumes due to their performance characteristics, which do not support the rapid read/write operations required for operating system files.

Incorrect Options:

1. **General Purpose SSD (gp2)**
2. **Provisioned IOPS SSD (io1)**
3. **Instance Store**

These options can all be used as boot volumes. General Purpose SSDs provide a balance of price and performance for a wide variety of transactional applications. Provisioned IOPS SSDs are designed for I/O-intensive applications such as large relational or NoSQL databases. Instance Store provides temporary block-level storage located on disks that are physically attached to the host server.

Conclusion / Important Points to Memorize:

- Know which EBS volume types can and cannot be used as boot volumes.
- SSD-based volumes and Instance Stores are suitable for boot volumes, while certain HDD-based volumes are not.

Question 4:

The payroll department at a company initiates several computationally intensive workloads on Amazon EC2 instances at a designated hour on the last day of every month. The department has noticed a trend of severe performance lag during this hour. The engineering team has suggested using an Auto Scaling group to ensure 10 Amazon EC2 instances are available during this peak usage hour. For normal operations, only 2 Amazon EC2 instances are enough.

Correct Option:

Configure your Auto Scaling group by creating a scheduled action that kicks off at the designated hour on the last day of the month. Set the desired capacity of instances to 10.

Explanation and Context:

Scheduled actions in Auto Scaling groups allow you to automatically increase or decrease the number of instances based on predictable load changes. Setting the desired capacity to 10 just before the peak period ensures that enough resources are available to handle the load without manual intervention.

Incorrect Options:

1. **Configure a target tracking policy**: This would not guarantee that exactly 10 instances are running at the required time.
2. **Use a simple tracking policy**: Similar to the target tracking policy, this does not ensure the precise number of needed instances.
3. **Set the min and max counts to 10**: This configuration would permanently limit the group to 10 instances, which is not cost-effective outside the peak period.

Conclusion / Important Points to Memorize:

- Use scheduled actions for predictable, cyclical changes in instance demand within an Auto Scaling group.
  - Understand the differences between various scaling policies and their appropriate applications.

Question 5:

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.

Correct Option:

Amazon API Gateway, Amazon Simple Queue Service (Amazon SQS), and Amazon Kinesis.

Explanation and Context:

- **Amazon API Gateway** can throttle requests to manage traffic spikes.
- **Amazon SQS** buffers requests, smoothing out sudden increases in traffic without losing messages.
- **Amazon Kinesis** deals with real-time data streams and can buffer incoming data.

Incorrect Options:

1. **Elastic Load Balancer**: This distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, but does not inherently throttle or buffer requests.
2. **AWS Lambda**: While Lambda can scale automatically in response to incoming requests, it does not buffer or throttle requests and may be overwhelmed by a sudden spike.

Conclusion / Important Points to Memorize:

- Utilize Amazon API Gateway for throttling, Amazon SQS for buffering messages, and Amazon Kinesis for managing data streams during traffic spikes.

Question 6:

An organization wants to delegate access to a set of users from the development environment so that they can access some resources in the production environment which is managed under another AWS account.

Correct Option:

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

Explanation and Context:

IAM roles provide a secure way to grant permissions to entities that you trust. By creating an IAM role with appropriate permissions and allowing users from the development environment to assume this role, you provide them with necessary access without the need to share credentials.

Incorrect Options:

1. **Create new IAM user credentials for the production environment**: This method involves sharing credentials, which is less secure and goes against AWS best practices.
2. **It is not possible to access cross-account resources**: This statement is incorrect; AWS provides several mechanisms for cross-account access, with IAM roles being one of the most secure and flexible options.

Conclusion / Important Points to Memorize:

- Use IAM roles for secure, cross-account access without the need for sharing user credentials.

Question 7:

A media company runs a photo-sharing web application accessed across three different countries. The application is deployed on Amazon Elastic Compute Cloud (Amazon EC2) instances running behind an Application Load Balancer. New government regulations require the company to block access from two countries and allow access only from the home country of the company.

Correct Option:

Configure AWS Web Application Firewall (AWS WAF) on the Application Load Balancer in an Amazon Virtual Private Cloud (Amazon VPC).

Explanation and Context:

AWS WAF allows you to control how traffic reaches your applications by enabling rules that allow or block traffic based on conditions such as IP addresses or geographic locations. This feature can effectively enforce compliance with geographic restrictions by blocking requests from specific countries.

Incorrect Options:

1. **Configure the security group for the Amazon EC2 instances**: Security groups act as a virtual firewall for instances to control inbound and outbound traffic, but they do not have the capability to filter traffic based on geographic location.
2. **Use Geo Restriction feature of Amazon CloudFront**: While CloudFront can restrict access based on geographic location, it is a content delivery network service, not a method for directly enforcing access restrictions at the load balancer level.
3. **Configure the security group on the Application Load Balancer**: Like EC2 security groups, ALB security groups cannot filter traffic based on geographic origins.

Conclusion / Important Points to Memorize:

- Use AWS WAF for implementing geographic restrictions directly on an Application Load Balancer.
- Understand the limitations of security groups and CloudFront Geo Restriction in controlling access based on geographic locations.

Question 8:

An Ivy League university is assisting NASA to find potential landing sites for unmanned missions to neighboring planets. The university uses High-Performance Computing (HPC) driven application architecture to identify these landing sites. Which EC2 instance topology should this application be deployed on?

Correct Option:

The Amazon 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.

Explanation and Context:

Cluster placement groups are ideal for HPC applications where low latency and high throughput are crucial. Instances in a cluster placement group are located physically close to each other in a single Availability Zone, which maximizes the bandwidth and minimizes the latency between instances.

Incorrect Options:

1. **Spread placement group**: This type of group ensures that instances are spread across underlying hardware to minimize correlated failures but does not optimize for network performance as needed for HPC.
2. **Partition placement group**: While it separates instances into different partitions across multiple Availability Zones, it does not provide the low-latency network performance required for tightly-coupled HPC applications.
3. **Auto Scaling group**: While important for maintaining availability and scaling, it doesn't specifically address the network performance needs of HPC applications.

Conclusion / Important Points to Memorize:

- Cluster placement groups are best suited for applications requiring high network throughput and low latency, such as HPC workloads.

Question 9:

A research group runs its flagship application on a fleet of Amazon EC2 instances for a specialized task that must deliver high random I/O performance. Each instance in the fleet has access to a dataset that is replicated across the instances by the application itself. Which option is the MOST cost-optimal and resource-efficient solution to build this fleet of Amazon EC2 instances?

Correct Option:

Use Instance Store based Amazon EC2 instances.

Explanation and Context:

Instance Store provides high I/O performance by offering temporary block-level storage located on disks physically attached to the host machine. This setup is optimal for applications requiring high random I/O due to the direct attachment of storage to the instance, providing faster access speeds.

Incorrect Options:

1. **Amazon Elastic Block Store (EBS) based EC2 instances**: While EBS offers persistence, using it for high I/O operations like this scenario can be more expensive and less efficient than Instance Stores.
2. **Amazon EC2 instances with Amazon EFS mount points**: EFS is designed for durability and shared access, not for high random I/O performance.
3. **Amazon EC2 instances with access to Amazon S3 based storage**: S3 is object storage that provides high durability but is not optimized for high-performance local disk access like Instance Store.

Conclusion / Important Points to Memorize:

- Instance Store is ideal for temporary storage needs where high I/O performance is crucial.

Question 10:

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 targets. The replicas have been assigned failover priority tiers. In the event of a failover, which read replica will Amazon Aurora promote?

Correct Option:

Tier-1 (32 terabytes)

Explanation and Context:

Amazon Aurora promotes the read replica with the highest priority tier (the lowest number) and, among replicas with the same tier, the largest size. In this scenario, the Tier-1 (32 terabytes) replica would be promoted because it has the highest priority and is the largest among those with the same priority.

Incorrect Options:

1. **Tier-15 (32 terabytes)**
2. **Tier-1 (16 terabytes)**
3. **Tier-10 (16 terabytes)**

These options are incorrect because they either have a lower priority or are smaller in size compared to the correct option.

Conclusion / Important Points to Memorize:

- Understand Aurora's failover process, particularly how priority tiers and instance sizes affect which replica is promoted during failover.

Question 11:

A company manages a multi-tier social media application running on Amazon EC2 instances behind an Application Load Balancer. The instances run in an Amazon EC2 Auto Scaling group across multiple Availability Zones and use an Amazon Aurora database. You have been tasked to make the application more resilient to periodic spikes in request rates. Which solutions would you recommend?

Correct Selections:

1. Use Amazon Aurora Replica.
2. Use Amazon CloudFront distribution in front of the Application Load Balancer.

Explanation and Context:

- **Amazon Aurora Replica**: This helps to scale read operations across multiple replicas to manage increased load without affecting the primary database's performance.
- **Amazon CloudFront**: Setting up a CDN in front of the application can cache content and reduce the load on the backend servers by serving requests from edge locations.

Incorrect Options:

1. **Use AWS Global Accelerator**: While it improves application availability and performance, it is not specifically designed for handling spikes in web application traffic.
2. **Use AWS Direct Connect and AWS Shield**: These services improve connectivity and protect against DDoS attacks, respectively, but do not directly address traffic spikes in application performance.

Conclusion / Important Points to Memorize:

- Utilizing Aurora Replicas and CloudFront can significantly enhance the resilience of a web application to handle traffic spikes by distributing the load and reducing direct hits to the backend infrastructure.

Question 12:

Amazon 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)

Correct Options:

1. **Dynamic content, as determined at request time (cache-behavior configured to forward all headers)**
2. **Proxy methods PUT/POST/PATCH/OPTIONS/DELETE go directly to the origin**

Explanation and Context:

- **Dynamic content**: When CloudFront is configured to forward all headers to the origin, it bypasses the cache because the content is expected to be unique per request, thus requiring direct origin fetches.
- **Proxy methods**: Methods like PUT, POST, PATCH, OPTIONS, and DELETE are typically used to modify data on the server, which necessitates direct interaction with the origin without caching.

Incorrect Options:

- **E-commerce assets such as product photos**
- **Static content such as style sheets, JavaScript files**
- **User-generated videos**

These content types are generally suitable for caching as they do not require real-time origin fetches unless specifically configured otherwise due to their static or less frequently changed nature.

Conclusion / Important Points to Memorize:

- Understanding which types of requests and content bypass CloudFront's cache can help optimize cache behavior settings for performance and cost-efficiency.

Question 13:

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

Correct Option:

Disable the service in the general settings

Explanation and Context:

Disabling Amazon GuardDuty in the general settings will cease all operations, and importantly, it will delete all findings and related data, adhering to the company's requirement for data deletion.

Incorrect Options:

- **Suspend the service in the general settings**: Suspending the service stops it from analyzing data but does not delete existing findings.
- **De-register the service under services tab**: There's no such option to de-register; disabling is the correct terminology and action.
- **Raise a service request with Amazon to completely delete the data from all their backups**: Unnecessary as disabling the service will automatically manage the deletion of data.

Conclusion / Important Points to Memorize:

- Properly disabling AWS services ensures compliance with data retention policies by removing potentially sensitive data.

Question 14:

A new DevOps engineer has joined a development team and wants to understand the replication capabilities for Amazon RDS Multi-AZ deployment as well as Amazon RDS Read-replicas. Which of the following correctly summarizes these capabilities?

Correct Option:

Multi-AZ follows synchronous replication and spans at least two Availability Zones (AZs) within a single region. Read replicas follow asynchronous replication and can be within an AZ, Cross-AZ, or Cross-Region.

Explanation and Context:

- **Multi-AZ**: Synchronous replication ensures that a primary DB instance and its secondary (standby) instance are always in sync. This setup provides high availability and data durability.
- **Read replicas**: Use asynchronous replication, allowing for scaling read capacity by replicating changes to the replica after they occur on the primary.

Incorrect Options:

- Other options incorrectly describe the replication methods or configurations between Multi-AZ and Read replicas, confusing their specific roles and technical behavior.

Conclusion / Important Points to Memorize:

- Know the distinction between Multi-AZ (for high availability) and Read replicas (for read scalability) in Amazon RDS.

Question 15

The DevOps team at an e-commerce company wants to perform some maintenance work on a specific Amazon 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)

Correct Options:

1. 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.
2. 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.

Explanation:

1. **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**: You can put an instance into the Standby state to update software or troubleshoot it. Instances in Standby are part of the Auto Scaling group but do not actively handle application traffic. This allows maintenance without triggering replacement.
2. **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 marked as unhealthy and creates new ones. By suspending this process, you prevent automatic replacement during maintenance.

Incorrect Options:

- **Take a snapshot of the instance, create a new Amazon Machine Image (AMI) and then launch a new instance using this AMI. Apply the maintenance patch to this new instance and then add it back to the Auto Scaling Group by using the manual scaling policy. Terminate the earlier instance that had the maintenance issue**: This approach is not time/resource efficient for applying a simple maintenance patch.
- **Delete the Auto Scaling group and apply the maintenance fix to the given instance. Create a new Auto Scaling group and add all the instances again using the manual scaling policy**: Deleting and recreating the Auto Scaling group is not recommended for applying a maintenance patch.
- **Suspend the ScheduledActions 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 ScheduledActions process type again**: This option is irrelevant to the given use case as it suspends scheduled scaling actions, not related to the immediate replacement issue during maintenance.

Conclusion / Concepts to Memorize:

- **Standby State**: Useful for temporarily removing an instance from service for maintenance without triggering replacement.
  - **ReplaceUnhealthy Process**: Should be suspended to prevent automatic replacement of instances during maintenance.

Question 16

A news network uses Amazon Simple Storage Service (Amazon S3) to aggregate the raw video footage from its reporting teams across the US. The news network has recently expanded into new geographies in Europe and Asia. The technical teams at the overseas branch offices have reported huge delays in uploading large video files to the destination Amazon S3 bucket.

Which of the following are the MOST cost-effective options to improve the file upload speed into Amazon S3 (Select two)?

Correct Options:

- Use Amazon S3 Transfer Acceleration (Amazon S3TA) to enable faster file uploads into the destination S3 bucket
- Use multipart uploads for faster file uploads into the destination Amazon S3 bucket

Explanation:

- **Use Amazon S3 Transfer Acceleration (Amazon S3TA) to enable faster file uploads into the destination S3 bucket**: Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket. It takes advantage of Amazon CloudFront's globally distributed edge locations. As the data arrives at an edge location, it is routed to Amazon S3 over an optimized network path.
- **Use multipart uploads for faster file uploads into the destination Amazon S3 bucket**: Multipart upload allows you to upload a single object as a set of parts. Each part is a contiguous portion of the object's data. You can upload these parts independently and in any order. If the transmission of any part fails, you can retransmit that part without affecting other parts. Multipart upload provides improved throughput, which facilitates faster file uploads, especially for large files.

Incorrect Options:

- **Create multiple AWS Direct Connect connections between the AWS Cloud and branch offices in Europe and Asia. Use the direct connect connections for faster file uploads into Amazon S3**: AWS Direct Connect is a dedicated network connection solution that takes significant time (several months) to provision and is an overkill for the given use-case.
- **Create multiple AWS Site-to-Site VPN connections between the AWS Cloud and branch offices in Europe and Asia. Use these VPN connections for faster file uploads into Amazon S3**: AWS Site-to-Site VPN provides secure connectivity but does not help in accelerating the file transfer speeds for the given use-case.
- **Use AWS Global Accelerator for faster file uploads into the destination Amazon S3 bucket**: AWS Global Accelerator improves the availability and performance of applications but does not specifically accelerate file transfers into Amazon S3.

Conclusion / Concepts to Memorize:

- **Amazon S3 Transfer Acceleration**: Useful for speeding up data transfers to S3 buckets from remote locations by using edge locations.
  - **Multipart Uploads**: Helps in faster and more reliable uploads of large objects to S3 by splitting the object into multiple parts and uploading them concurrently.

Question 17

The engineering team at a data analytics company has observed that its flagship application functions at its peak performance when the underlying Amazon Elastic Compute Cloud (Amazon EC2) instances have a CPU utilization of about 50%. The application is built on a fleet of Amazon EC2

instances managed under an Auto Scaling group. The workflow requests are handled by an internal Application Load Balancer that routes the requests to the instances.

As a solutions architect, what would you recommend so that the application runs near its peak performance state?

Correct Option:

Configure the Auto Scaling group to use target tracking policy and set the CPU utilization as the target metric with a target value of 50%

Explanation:

- **Configure the Auto Scaling group to use target tracking policy and set the CPU utilization as the target metric with a target value of 50%**: Target tracking scaling policies allow you to set a target value for a specific metric, and Amazon EC2 Auto Scaling automatically adjusts the number of instances to maintain the target metric. For example, you can configure a target tracking scaling policy to keep the average aggregate CPU utilization of your Auto Scaling group at 50%. This aligns with the requirement of maintaining the CPU utilization at 50% to achieve peak performance.

Incorrect Options:

- **Configure the Auto Scaling group to use step scaling policy and set the CPU utilization as the target metric with a target value of 50%**: Step scaling policies adjust the number of instances based on a series of thresholds. They are not as dynamic as target tracking policies and do not maintain a specific target value.
- **Configure the Auto Scaling group to use simple scaling policy and set the CPU utilization as the target metric with a target value of 50%**: Simple scaling policies trigger scaling activities based on a single threshold, which is less flexible compared to target tracking policies.
- **Configure the Auto Scaling group to use an Amazon CloudWatch alarm triggered on a CPU utilization threshold of 50%**: CloudWatch alarms can trigger scaling actions, but they do not provide the continuous adjustment and dynamic response that target tracking policies offer.

Conclusion / Concepts to Memorize:

- **Target Tracking Scaling Policy**: Ideal for maintaining a specific metric at a target value, automatically adjusting capacity as needed.
- **Step and Simple Scaling Policies**: Useful for predefined scaling actions based on specific conditions but less flexible than target tracking.
  - **CloudWatch Alarms**: Can trigger scaling actions but do not provide the continuous adjustment capabilities of target tracking policies.

Question 18

A junior scientist working with the Deep Space Research Laboratory at NASA is trying to upload a high-resolution image of a nebula into Amazon S3. The image size is approximately 3 gigabytes. The junior scientist is using Amazon S3 Transfer Acceleration (Amazon S3TA) for faster image upload. It turns out that Amazon S3TA did not result in an accelerated transfer.

Given this scenario, which of the following is correct regarding the charges for this image transfer?

Correct Option:

The junior scientist does not need to pay any transfer charges for the image upload.

Explanation:

- **The junior scientist does not need to pay any transfer charges for the image upload**: There are no Amazon S3 data transfer charges when data is transferred from the internet. With Amazon S3 Transfer Acceleration (S3TA), you only pay for transfers that are accelerated. Since the transfer was not accelerated, there are no additional charges for using S3TA.

Incorrect Options:

- **The junior scientist only needs to pay S3TA transfer charges for the image upload**: Since the transfer was not accelerated, there are no S3TA transfer charges to be paid.
- **The junior scientist only needs to pay Amazon S3 transfer charges for the image upload**: There are no S3 data transfer charges when data is transferred from the internet.
- **The junior scientist needs to pay both S3 transfer charges and S3TA transfer charges for the image upload**: There are no Amazon S3 data transfer charges for internet data transfers, and since S3TA did not result in an accelerated transfer, there are no S3TA transfer charges to be paid.

Conclusion / Concepts to Memorize:

- **Amazon S3 Transfer Acceleration (S3TA)**: You only pay for transfers that are accelerated. If the transfer is not accelerated, no additional charges apply.
- **Data Transfer Charges**: There are no Amazon S3 data transfer charges when data is transferred from the internet.
  - **AWS Cost Management**: Understanding how AWS services charge for data transfer can help manage and predict costs effectively.

Question 19

An Electronic Design Automation (EDA) application produces massive volumes of data that can be divided into two categories. The 'hot data' needs to be both processed and stored quickly in a parallel and distributed fashion. The 'cold data' needs to be kept for reference with quick access for reads and updates at a low cost.

Which of the following AWS services is BEST suited to accelerate the aforementioned chip design process?

Correct Option:

Amazon FSx for Lustre

Explanation

- **Amazon FSx for Lustre**: This service makes it easy and cost-effective to launch and run the world's most popular high-performance file system. It is designed for workloads that require fast storage to keep up with compute, making it ideal for 'hot data'. FSx for Lustre integrates with Amazon S3, allowing for easy storage and quick access to 'cold data' as well. Therefore, this option is the best fit for the given problem statement.
  - **Reference**: AWS Documentation on Amazon FSx for Lustre

Incorrect Options:

- **Amazon FSx for Windows File Server**: Provides fully managed, highly reliable file storage but does not allow S3 objects to be presented as files or write changed data back to S3, which is essential for handling 'cold data' efficiently.
- **Amazon EMR**: Industry-leading cloud big data platform for processing vast amounts of data but does not offer the same storage and processing speed as FSx for Lustre, making it less suitable for the high-performance workflow scenario.
- **AWS Glue**: Fully managed ETL service for preparing and loading data for analytics, but it does not offer the same storage and processing speed as FSx for Lustre, making it less ideal for the given use case.

Conclusion / Concepts to Memorize:

- **Amazon FSx for Lustre**: Best for high-performance computing (HPC) workloads requiring fast storage to keep up with compute, such as machine learning, financial modeling, and video processing. Integrates with Amazon S3 for managing 'cold data'.
- **Amazon FSx for Windows File Server**: Suitable for applications that need fully managed, highly reliable Windows file storage accessible over SMB.
- **Amazon EMR**: Ideal for big data processing using Hadoop, Spark, HBase, etc., but not optimized for the same high-performance storage needs as FSx for Lustre.
  - **AWS Glue**: Best for ETL jobs to prepare and load data for analytics but not for high-performance storage and processing.

Question 20

Which of the following features of an Amazon S3 bucket can only be suspended and not disabled once it has been enabled?

Correct Option:

Versioning

Explanation:

- **Versioning**: Once you version-enable a bucket, it can never return to an unversioned state. Versioning can only be suspended once it has been enabled.
  - **Reference**: AWS Documentation on Versioning

Incorrect Options:

- **Server Access Logging**: Can be enabled and disabled.
- **Static Website Hosting**: Can be enabled and disabled.
- **Requester Pays**: Can be enabled and disabled.

Question 21

The product team at a startup has figured out a market need to support both stateful and stateless client-server communications via the application programming interface (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 Amazon API Gateway. Which of the following would you identify as correct?

Correct Option:

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

Explanation:

- **Amazon API Gateway**:
  - **RESTful APIs**: Enable stateless client-server communication, utilizing standard HTTP methods like GET, POST, PUT, PATCH, and DELETE.

- **WebSocket APIs**: Adhere to the WebSocket protocol, enabling stateful, full-duplex communication between client and server, allowing real-time two-way communication.

Incorrect Options:

1. **Stateful RESTful APIs**:
   Amazon API Gateway creates RESTful APIs that enable stateful client-server communication and Amazon API Gateway also creates WebSocket APIs that adhere to the WebSocket protocol, which enables stateful, full-duplex communication between client and server.
2. **Stateless WebSocket APIs**:
   Amazon API Gateway creates RESTful APIs that enable stateless client-server communication and Amazon API Gateway also creates WebSocket APIs that adhere to the WebSocket protocol, which enables stateless, full-duplex communication between client and server.
3. **Stateful RESTful APIs and Stateless WebSocket APIs**:
   Amazon API Gateway creates RESTful APIs that enable stateful client-server communication and Amazon API Gateway also creates WebSocket APIs that adhere to the WebSocket protocol, which enables stateless, full-duplex communication between client and server

Summary:

Amazon API Gateway supports stateless RESTful APIs and stateful WebSocket APIs, making the correct option to accurately describe the capabilities and distinctions between these two types of APIs.

Question 22

A development team requires permissions to list an Amazon S3 bucket and delete objects from that bucket. A systems administrator has created the following IAM policy to provide access to the bucket and applied that policy to the group. The group is not able to delete objects in the bucket. The company follows the principle of least privilege.

Given Policy:	**Correct Option**:

"Version": "2021-10-17",
"Statement": [
{
"Action": [
"s3:ListBucket",
"s3:DeleteObject"
],
"Resource": [
"arn:aws:s3:::example-bucket"
],
"Effect": "Allow"
}
]
	{
	"Action": [
	"s3:DeleteObject"
	],
	"Resource": [
	"arn:aws:s3:::example-bucket/*"
	],
	"Effect": "Allow"
	}

Explanation:

- The existing policy allows listing the bucket but does not correctly permit deleting objects because the s3:DeleteObject action must be specified on the objects within the bucket, not just the bucket itself.
- The correct statement adds the necessary permissions to delete objects within the example-bucket.

Incorrect Options:

Invalid action value	Violates least privilege principle	Incorrect resource name
{ "Action": ["s3:*Object"], "Resource": ["arn:aws:s3:::example-bucket/*"],	{ "Action": ["s3:*"], "Resource": ["arn:aws:s3:::example-bucket/*"], "Effect": "Allow"	{ "Action": ["s3:DeleteObject"], "Resource": ["arn:aws:s3:::example-bucket*"],

"Effect": "Allow"	}	"Effect": "Allow"
}		}

Summary:

The correct option ensures that the necessary delete permissions are granted on the objects within the S3 bucket, following the principle of least privilege by only allowing the specific actions required.

Question 23

A geological research agency maintains the seismological data for the last 100 years. The data has a velocity of 1GB per minute. You would like to store the data with only the most relevant attributes to build a predictive model for earthquakes.

What AWS services would you use to build the most cost-effective solution with the LEAST amount of infrastructure maintenance?

Correct Option:

Ingest the data in Amazon Kinesis Data Firehose and use an intermediary AWS Lambda function to filter and transform the incoming stream before the output is dumped on Amazon S3

Explanation Behind Correct Option:

Amazon Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools. It can capture, transform, and load streaming data into Amazon S3, enabling near real-time analytics with existing business intelligence tools. It is a fully managed service that scales automatically, requires no ongoing administration, and can batch, compress, and encrypt the data before loading it, minimizing the amount of storage used and increasing security.

Incorrect Options and Their Brief Explanation:

1. **Ingest the data in Amazon Kinesis Data Analytics and use SQL queries to filter and transform the data before writing to Amazon S3**:

Kinesis Data Analytics cannot directly ingest data from the source as it ingests data either from Kinesis Data Streams or Kinesis Data Firehose, making it less efficient for this use case.

2. **Ingest the data in Amazon Kinesis Data Streams and use an intermediary AWS Lambda function to filter and transform the incoming stream before the output is dumped on Amazon S3**:

Kinesis Data Streams does not offer a ready-made integration to reliably dump data into Amazon S3. It requires custom coding to maintain the buffer reliably and ensure no data is lost.

3. **Ingest the data in a Spark Streaming Cluster on Amazon EMR and use Spark Streaming transformations before writing to Amazon S3**:

Using an EMR cluster would require managing the underlying infrastructure, which contradicts the requirement for the least amount of infrastructure maintenance.

Conclusion / Points to Memorize:

Amazon Kinesis Data Firehose is ideal for capturing, transforming, and loading streaming data into AWS data stores such as Amazon S3. It provides a serverless, fully managed, and scalable solution, minimizing infrastructure maintenance.

Question 24

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 70 Terabytes 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)

Correct options:

1. Order 10 AWS Snowball Edge Storage Optimized devices to complete the one-time data transfer
2. Setup AWS Site-to-Site VPN to establish on-going connectivity between the on-premises data center and AWS Cloud

Explanation behind correct options:

- **Order 10 AWS Snowball Edge Storage Optimized devices to complete the one-time data transfer:**
  - **Reason:** AWS Snowball Edge Storage Optimized devices provide up to 80 Terabytes of usable storage, which makes them suitable for transferring large volumes of data securely and quickly. With a data footprint of 70 Terabytes per application and 10 applications in total, ordering 10 Snowball Edge devices ensures that each device can handle the data transfer requirements for each application within the two-week timeline.

  - **Details:** These devices also offer 40 vCPUs, 1 TB of SATA SSD storage, and up to 40 Gigabytes network connectivity, making them ideal for large-scale data transfer and pre-processing use cases.
- **Setup AWS Site-to-Site VPN to establish on-going connectivity between the on-premises data center and AWS Cloud:**
  - **Reason:** AWS Site-to-Site VPN provides secure, encrypted connectivity between the on-premises data center and AWS Cloud over the Internet. This solution is cost-effective and can be configured quickly to meet the immediate connectivity needs within the given timeframe.
  - **Details:** It utilizes IPSec to establish the connection, making it suitable for low to modest bandwidth requirements with the flexibility to handle the inherent variability of Internet-based connectivity.

Incorrect options and their explanations:

1. **Order 1 AWS Snowmobile to complete the one-time data transfer:**
   - **Reason:** AWS Snowmobile is designed for extremely large data transfers, up to 100 petabytes. It is not suitable for smaller datasets like the one described (700 Terabytes total), as it would be an overkill and not cost-effective for this use case.
   - **Details:** Snowmobile is best used for data transfers exceeding 10 petabytes in a single location.
2. **Setup AWS Direct Connect to establish connectivity between the on-premises data center and AWS Cloud:**
   - **Reason:** AWS Direct Connect involves setting up a dedicated network connection that requires significant monetary investment and takes more time (at least a month) to establish.
   - **Details:** This option is not feasible given the two-week timeline and the need for a quick and cost-effective solution.
3. **Order 70 AWS Snowball Edge Storage Optimized devices to complete the one-time data transfer:**
   - **Reason:** Ordering 70 devices is unnecessary and not cost-effective, as the data transfer can be completed with just 10 devices.
   - **Details:** Over-provisioning would result in unnecessary costs without providing additional benefits for this use case.

Conclusion / Points to Memorize:

- **AWS Snowball Edge Storage Optimized devices** are ideal for securely and efficiently transferring large volumes of data to AWS.
- **AWS Site-to-Site VPN** provides a quick and cost-effective way to establish secure connectivity between on-premises data centers and AWS Cloud.
  - **AWS Direct Connect** and **AWS Snowmobile** are not suitable for scenarios requiring quick setup or for smaller data volumes.

Question 25

As part of a pilot program, a biotechnology company wants to integrate data files from its on-premises analytical application with AWS Cloud via an NFS interface. Which of the following AWS services is the MOST efficient solution for the given use-case?

Correct option:

- AWS Storage Gateway - File Gateway

Explanation behind correct option:

- **AWS Storage Gateway - File Gateway:** AWS Storage Gateway is a hybrid cloud storage service that provides on-premises access to virtually unlimited cloud storage. The File Gateway offers a seamless way to connect on-premises applications to cloud storage via SMB or NFS-based access. It stores application data files and backup images as durable objects in Amazon S3, and it provides local caching for low-latency access. Since the company needs to integrate data files via an NFS interface, AWS Storage Gateway - File Gateway is the most efficient solution.

Incorrect options and their explanations:

1. **AWS Storage Gateway - Volume Gateway:**
   - Volume Gateway presents cloud-based iSCSI block storage volumes to on-premises applications. It does not support the NFS interface required by the company, so this option is not correct.
2. **AWS Storage Gateway - Tape Gateway:**
   - Tape Gateway is used for moving tape backups to the cloud. It also does not support the NFS interface, making it an unsuitable choice for this use-case.
3. **AWS Site-to-Site VPN:**
   - AWS Site-to-Site VPN enables secure connections between on-premises networks and Amazon VPC using IPSec communication. However, it does not facilitate integration of data files via the NFS interface, so this option is not applicable.

Conclusion / Points to Memorize:

- **File Gateway for NFS Interface:** When integrating on-premises data with AWS Cloud via NFS interface, AWS Storage Gateway - File Gateway is the best solution.
- **Gateway Types:** Understand the differences among AWS Storage Gateway types: File Gateway (NFS/SMB), Volume Gateway (iSCSI block storage), and Tape Gateway (tape backups).

- **AWS Site-to-Site VPN:** This service is for secure network connections, not for file integration via NFS.

Question 26

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?

Correct option:

- Consolidated billing has not been enabled. All the AWS accounts should fall under a single consolidated billing for the monthly fee to be charged only once

Explanation behind correct option:

- **Consolidated Billing:** If your organization has multiple AWS accounts, you can subscribe multiple AWS Accounts to AWS Shield Advanced by individually enabling it on each account using the AWS Management Console or API. You will pay the monthly fee once as long as the AWS accounts are all under a single consolidated billing, and you own all the AWS accounts and resources in those accounts. Without consolidated billing, each account would be charged the AWS Shield Advanced monthly fee, leading to higher costs.

Incorrect options and their explanations:

1. **AWS Shield Advanced is being used for custom servers, that are not part of AWS Cloud, thereby resulting in increased costs:**
   - AWS Shield Advanced does offer protection to resources outside of AWS. However, this should not cause an unexpected spike in billing costs directly related to AWS Shield Advanced services for AWS resources.
2. **AWS Shield Advanced also covers AWS Shield Standard plan, thereby resulting in increased costs:**
   - AWS Shield Standard is automatically enabled for all AWS customers at no additional cost. AWS Shield Advanced is an optional paid service and its cost does not include any additional charges for the AWS Shield Standard plan, which is free.
3. **Savings Plans has not been enabled for the AWS Shield Advanced service across all the AWS accounts:**
   - Savings Plans is a flexible pricing model that offers low prices on Amazon EC2 instances, AWS Lambda, and AWS Fargate usage, in exchange for a commitment to a consistent amount of usage (measured in $/hour) for a 1 or 3 year term. Savings Plans is not applicable for the AWS Shield Advanced service. This option is irrelevant to the issue of unexpectedly high costs.

Conclusion / Points to Memorize:

- **Consolidated Billing:** Ensure all AWS accounts are under a single consolidated billing to avoid multiple charges for services like AWS Shield Advanced.
- **AWS Shield Advanced Costs:** Understand that AWS Shield Advanced charges a monthly fee per account, which can be reduced by using consolidated billing.
- **Irrelevant Cost Factors:** Recognize that AWS Shield Standard is free and not part of the cost issues for AWS Shield Advanced, and Savings Plans do not apply to AWS Shield Advanced.

Question 27

One of the biggest football leagues in Europe has granted the distribution rights for live streaming its matches in the USA to a Silicon Valley-based streaming services company. As per the terms of distribution, the company must make sure that only users from the USA are able to live stream the matches on their platform. Users from other countries in the world must be denied access to these live-streamed matches. Which of the following options would allow the company to enforce these streaming restrictions? (Select two)

Correct options:

- Use Amazon Route 53 based geolocation routing policy to restrict the distribution of content to only the locations in which you have distribution rights.
- Use georestriction to prevent users in specific geographic locations from accessing content that you're distributing through an Amazon CloudFront web distribution.

Explanation behind correct options:

1. **Amazon Route 53 Geolocation Routing Policy:**
   - Geolocation routing lets you choose the resources that serve your traffic based on the geographic location of your users, meaning the location that DNS queries originate from. This can be used to ensure that only users from the USA can access the live-streamed matches, complying with the distribution rights.
2. **Amazon CloudFront Georestriction:**

- Georestriction, also known as geo-blocking, can be used to prevent users in specific geographic locations from accessing content distributed through an Amazon CloudFront web distribution. This ensures that only users from allowed regions (e.g., the USA) can access the content.

Incorrect options and their explanations:

1. **Amazon Route 53 Latency-Based Routing Policy:**
   - Latency-based routing is used to route traffic to the region that provides the best latency. It does not provide a mechanism to restrict access based on geographic location.
2. **Amazon Route 53 Weighted Routing Policy:**
   - Weighted routing allows you to route traffic to multiple resources in proportions that you specify. It is typically used for load balancing and does not support restricting access based on geographic location.
3. **Amazon Route 53 Failover Routing Policy:**
   - Failover routing is used to route traffic to a resource when the resource is healthy and to a different resource when the first resource is unhealthy. It does not offer geolocation-based restrictions.

Conclusion / Points to Memorize:

- **Geolocation Routing (Route 53):** Best for routing traffic based on the geographic location of users.
- **Georestriction (CloudFront):** Best for restricting access to content based on geographic locations.
- **Latency-Based Routing, Weighted Routing, and Failover Routing (Route 53):** Not suitable for restricting access based on geographic locations.

Question 28

A gaming company is looking at improving the availability and performance of its global flagship application which utilizes User Datagram Protocol and needs to support fast regional failover in case an AWS Region goes down. The company wants to continue using its own custom Domain Name System (DNS) service. Which of the following AWS services represents the best solution for this use-case?

Correct option:

AWS Global Accelerator

Explanation behind correct option:

- **AWS Global Accelerator:** Utilizes the Amazon global network to improve the performance of applications by lowering first-byte latency and jitter and increasing throughput compared to the public internet. It supports fast regional failover and is suitable for non-HTTP use cases like gaming (UDP), IoT (MQTT), or Voice over IP, as well as HTTP use cases requiring static IP addresses or fast regional failover. It does not require a change in the DNS service, making it the best fit for the use case.

Incorrect options and their explanations:

1. **Amazon CloudFront:** A content delivery network (CDN) service that delivers data, videos, applications, and APIs globally with low latency and high transfer speeds. While CloudFront improves performance for cacheable and dynamic content, it does not specifically address the need for UDP support or fast regional failover as required by the use case.
2. **AWS Elastic Load Balancing (ELB):** Provides load balancing within one region, distributing incoming application traffic across backends like Amazon EC2 instances or ECS tasks. ELB does not offer traffic management across multiple regions, which is essential for the global application requiring fast regional failover.
3. **Amazon Route 53:** A highly available and scalable cloud DNS web service designed to route end users to Internet applications by translating domain names into IP addresses. However, the use case specifies that the company wants to continue using its own custom DNS service, ruling out Route 53.

Conclusion / Points to Memorize:

- **AWS Global Accelerator:** Best for improving performance and availability for applications using TCP or UDP, supports fast regional failover, and does not require changes to existing DNS services.
- **Amazon CloudFront:** Best for delivering content with low latency and high transfer speeds but not ideal for non-HTTP use cases requiring UDP support.
- **AWS Elastic Load Balancing (ELB):** Suitable for load balancing within a single region but not for global traffic management across multiple regions.
- **Amazon Route 53:** Ideal for DNS routing but not applicable if a custom DNS service is in use.

Question 29

A technology blogger wants to write a review on the comparative pricing for various storage types available on AWS Cloud. The blogger has created a test file of size 1 gigabyte with some random data. Next, he copies this test file into AWS S3 Standard storage class, provisions an Amazon EBS volume (General Purpose SSD (gp2)) with 100 gigabytes of provisioned storage and copies the test file into the Amazon EBS volume, and lastly copies the test file into an Amazon EFS Standard Storage filesystem. At the end of the month, he analyses the bill for costs incurred on the respective storage types for the test file. What is the correct order of the storage charges incurred for the test file on these three storage types?

Correct option:

Cost of test file storage on Amazon S3 Standard < Cost of test file storage on Amazon EFS < Cost of test file storage on Amazon EBS

Explanation behind correct option:

- **Amazon S3 Standard Storage Cost:** The pricing for Amazon S3 Standard storage is $0.023 per GB per month. Therefore, the monthly storage cost on S3 for the 1 GB test file is $0.023.
- **Amazon EFS Standard Storage Cost:** The pricing for Amazon EFS Standard Storage is $0.30 per GB per month. Therefore, the cost for storing the test file on EFS is $0.30 for the month.
- **Amazon EBS General Purpose SSD (gp2) Cost:** The charges for Amazon EBS General Purpose SSD (gp2) volumes are $0.10 per GB-month of provisioned storage. Therefore, for a provisioned storage of 100 GB for this use case, the monthly cost on EBS is $0.10 * 100 = $10. This cost is irrespective of how much storage is actually consumed by the test file.

Thus, the correct order of costs is:

1. **Amazon S3 Standard**: $0.023
2. **Amazon EFS**: $0.30
3. **Amazon EBS**: $10

Incorrect options and their explanations:

1. **Cost of test file storage on Amazon S3 Standard < Cost of test file storage on Amazon EBS < Cost of test file storage on Amazon EFS**
   - This order is incorrect because the cost of storing the test file on Amazon EBS is higher than on Amazon EFS.
2. **Cost of test file storage on Amazon EFS < Cost of test file storage on Amazon S3 Standard < Cost of test file storage on Amazon EBS**
   - This order is incorrect because the cost of storing the test file on Amazon EFS is higher than on Amazon S3 Standard.
3. **Cost of test file storage on Amazon EBS < Cost of test file storage on Amazon S3 Standard < Cost of test file storage on Amazon EFS**
   - This order is incorrect because the cost of storing the test file on Amazon EBS is the highest among the three storage types.

Conclusion / Points to Memorize:

- **Amazon S3 Standard:** $0.023 per GB per month
- **Amazon EFS:** $0.30 per GB per month
- **Amazon EBS (gp2):** $0.10 per GB per month of provisioned storage (100 GB minimum, resulting in $10 for this use case)
  - **Order of Cost Efficiency:** S3 Standard < EFS < EBS

Question 30

The development team at an e-commerce startup has set up multiple microservices running on Amazon EC2 instances under an Application Load Balancer. The team wants to route traffic to multiple back-end services based on the URL path of the HTTP header. So it wants requests for https://www.example.com/orders to go to a specific microservice and requests for https://www.example.com/products to go to another microservice. Which of the following features of Application Load Balancers can be used for this use-case?

Correct option:

Path-based Routing

Explanation behind correct option:

- **Path-based Routing:** This feature allows routing of client requests based on the URL path in the HTTP header. It is ideal for directing different types of traffic to specific microservices. For instance, requests to /orders can be routed to one microservice, while requests to /products can be directed to another.

Path-based Routing Overview:

- **Path Conditions:** Used to define rules that route requests based on the URL in the request.
- **Pattern Application:** The path pattern is applied only to the path of the URL, not to its query parameters.

Incorrect options and their explanations:

1. **Query string parameter-based routing:** This type routes requests based on query parameters in the URL, not suitable for routing based on the path.

2. **HTTP header-based routing:** This type routes requests based on values in HTTP headers, not the URL path.
3. **Host-based Routing:** This type routes requests based on the Host field in the HTTP header, suitable for routing different domains, not paths within the same domain.

Conclusion / Points to Memorize:

- **Path-based Routing:** Ideal for routing traffic based on specific URL paths.
    - **Query string, HTTP header, and Host-based Routing:** Not suitable for this use-case as they do not route based on the URL path.

Question 31

The engineering team at an in-home fitness company is evaluating multiple in-memory data stores with the ability to power its on-demand, live leaderboard. The company's leaderboard requires high availability, low latency, and real-time processing to deliver customizable user data for the community of users working out together virtually from the comfort of their home. As a solutions architect, which of the following solutions would you recommend? (Select two)

Correct options:

1. Power the on-demand, live leaderboard using Amazon ElastiCache for Redis as it meets the in-memory, high availability, low latency requirements
2. Power the on-demand, live leaderboard using Amazon DynamoDB with DynamoDB Accelerator (DAX) as it meets the in-memory, high availability, low latency requirements

Explanation behind correct options:

1. **Amazon ElastiCache for Redis:**
    - **Explanation:** Amazon ElastiCache for Redis is a blazing fast in-memory data store that provides sub-millisecond latency to power internet-scale real-time applications. It is suitable for real-time transactional and analytical processing use cases such as caching, chat/messaging, gaming leaderboards, geospatial, machine learning, media streaming, queues, real-time analytics, and session store. Therefore, it meets the requirements for high availability, low latency, and real-time processing, making it an ideal choice for powering the live leaderboard.
    - **Amazon ElastiCache for Redis Overview:**
        - Internet-scale applications
        - Use cases include real-time transactions, chat, BI and analytics, session store, gaming leaderboards, and cache
2. **Amazon DynamoDB with DynamoDB Accelerator (DAX):**
    - **Explanation:** Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It is fully managed, multiregion, multimaster, durable, and comes with built-in security, backup and restore, and in-memory caching for internet-scale applications. DAX is a DynamoDB-compatible caching service that enables fast in-memory performance for demanding applications. Hence, DynamoDB with DAX can be effectively used to power the live leaderboard, fulfilling the requirements of in-memory, high availability, and low latency.

Incorrect options and their explanations:

1. **Amazon Neptune:**
    - **Explanation:** Amazon Neptune is a fast, reliable, fully-managed graph database service designed for applications that work with highly connected datasets. It is not an in-memory database, so it does not meet the requirement for in-memory data storage and processing.
2. **Amazon DynamoDB:**
    - **Explanation:** While Amazon DynamoDB is a highly available and low latency database, it is not an in-memory database by default. Without the use of DynamoDB Accelerator (DAX), it would not fulfill the in-memory requirement.
3. **Amazon RDS for Aurora:**
    - **Explanation:** Amazon Aurora is a MySQL and PostgreSQL-compatible relational database built for the cloud, offering high performance and availability. However, it is not an in-memory database, making it unsuitable for this particular use case that demands in-memory processing.

Conclusion / Points to Memorize:

- **Amazon ElastiCache for Redis:** Suitable for real-time applications requiring sub-millisecond latency, in-memory data storage, high availability, and low latency.
- **Amazon DynamoDB with DAX:** Suitable for real-time applications requiring single-digit millisecond performance, in-memory caching, high availability, and low latency.
- **Amazon Neptune, DynamoDB (without DAX), and Amazon RDS for Aurora:** Not suitable for this use case due to lack of in-memory processing capabilities.

Question 32

A telecom company operates thousands of hardware devices like switches, routers, cables, etc. The real-time status data for these devices must be fed into a communications application for notifications. Simultaneously, another analytics application needs to read the same real-time status data and analyze all the connecting lines that may go down because of any device failures. As an AWS Certified Solutions Architect – Associate, which of the following solutions would you suggest, so that both the applications can consume the real-time status data concurrently?

Correct option:

Amazon Kinesis Data Streams

Explanation behind correct option:

- **Amazon Kinesis Data Streams:** Kinesis Data Streams enables real-time processing of streaming big data. It provides ordering of records and the ability to read and/or replay records in the same order to multiple Amazon Kinesis Applications. The Amazon Kinesis Client Library (KCL) delivers all records for a given partition key to the same record processor, making it easier to build multiple applications reading from the same Amazon Kinesis data stream concurrently. This is ideal for the telecom company's requirement where both the communications and analytics applications need to consume the same real-time status data concurrently.

AWS recommends Amazon Kinesis Data Streams for use cases with the following requirements:

1. Routing related records to the same record processor (as in streaming MapReduce), which simplifies counting and aggregation.
2. Ordering of records, ensuring data transfer maintains the order of log statements.
3. Allowing multiple applications to consume the same stream concurrently, such as updating a real-time dashboard and archiving data to Amazon Redshift.
4. The ability to consume records in the same order a few hours later, suitable for applications like billing and auditing that run at different times but need the same data.

Incorrect options and their explanations:

1. **Amazon Simple Notification Service (SNS):**
   - **Explanation:** Amazon SNS is a pub/sub messaging service for high-throughput, push-based, many-to-many messaging. It is designed for notification purposes and does not support real-time data processing required by the given use case.
2. **Amazon Simple Queue Service (SQS) with Amazon Simple Notification Service (SNS):**
   - **Explanation:** While Amazon SQS offers a reliable, scalable queue for storing messages, it is not suitable for real-time concurrent data consumption by multiple applications. The combination of SQS with SNS does not provide the same capabilities for concurrent stream processing and real-time data handling as Kinesis Data Streams.
3. **Amazon Simple Queue Service (SQS) with Amazon Simple Email Service (SES):**
   - **Explanation:** Amazon SQS and Amazon SES do not fit the use case requirements for real-time processing and concurrent consumption of data streams. SES is primarily used for email services and does not handle real-time data streams.

Conclusion / Points to Memorize:

- **Amazon Kinesis Data Streams:** Ideal for real-time processing of streaming data, supports multiple consumers, and maintains data order.
- **Amazon SNS:** Used for notifications, not suitable for real-time processing.
- **Amazon SQS:** Provides message queuing, but not ideal for concurrent real-time data consumption.
- **Amazon SES:** Email service, not relevant for real-time data processing.

Understanding the capabilities of different AWS services is crucial in designing the most efficient and cost-effective solutions for specific use cases.

Question 33

The solo founder at a tech startup has just created a brand new AWS account. The founder has provisioned an Amazon EC2 instance 1A which is running in AWS Region A. Later, he takes a snapshot of the instance 1A and then creates a new Amazon Machine Image (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?

Correct option:

1 Amazon EC2 instance, 1 AMI and 1 snapshot exist in Region B

Explanation behind correct option:

- **Amazon Machine Image (AMI):** An AMI provides the information required to 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. Therefore, in Region B, there will be 1 Amazon EC2 instance, 1 AMI, and 1 snapshot.

Incorrect options and their explanations:

1. **1 Amazon EC2 instance and 1 AMI exist in Region B:**
   - **Explanation:** This option is incorrect because it does not account for the snapshot that is created when the AMI is copied to Region B.
2. **1 Amazon EC2 instance and 1 snapshot exist in Region B:**
   - **Explanation:** This option is incorrect because it does not account for the AMI that is created when the snapshot is used to provision an instance in Region B.
3. **1 Amazon EC2 instance and 2 AMIs exist in Region B:**
   - **Explanation:** This option is incorrect because there is only one AMI created in Region B, not two. The additional AMI mentioned in this option does not exist.

Conclusion / Points to Memorize:

- **AMI Lifecycle:** An AMI includes one or more EBS snapshots, launch permissions, and block device mappings.
- **Snapshot Creation:** When an AMI is copied to a new region, a snapshot is automatically created in that region.
- **Entities in Region:** After copying an AMI and provisioning an instance in a new region, you will have 1 EC2 instance, 1 AMI, and 1 snapshot in the new region.
- **AWS Regions:** AMIs and their underlying snapshots must be explicitly copied to other regions to be used for instance launches in those regions.

Question 34

A file-hosting service uses Amazon Simple Storage Service (Amazon S3) under the hood to power its storage offerings. Currently, all the customer files are uploaded directly under a single Amazon 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?

Correct option:

Change the application architecture to create customer-specific custom prefixes within the single Amazon S3 bucket and then upload the daily files into those prefixed locations.

Explanation behind correct option:

- **Amazon S3 Performance Optimization:** Amazon S3 automatically scales to high request rates. For example, your application 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 and writes. Creating customer-specific custom prefixes within the single Amazon S3 bucket and then uploading the daily files into those prefixed locations will distribute the load, allowing higher throughput and avoiding request throttling.

Incorrect options and their explanations:

1. **Change the application architecture to create a new Amazon S3 bucket for each customer and then upload each customer's files directly under the respective buckets:**
   - **Explanation:** Creating a new Amazon S3 bucket for each new customer is inefficient in terms of resource availability because S3 buckets need to be globally unique. This approach also leads to potential management and scalability issues without significantly improving performance.
2. **Change the application architecture to create a new Amazon S3 bucket for each day's data and then upload the daily files directly under that day's bucket:**
   - **Explanation:** Creating a new Amazon S3 bucket for each day's data is inefficient as S3 bucket names need to be globally unique, which can lead to naming conflicts and management complexity. This approach is unnecessary because performance can be optimized using prefixes.
3. **Change the application architecture to use Amazon Elastic File System (Amazon EFS) instead of Amazon S3 for storing the customers' uploaded files:**
   - **Explanation:** Amazon EFS is designed for different use cases and is typically more expensive than S3. It is not a cost-optimal solution for storing large amounts of customer files that can be efficiently managed and scaled using S3 with proper prefixing strategies.

Conclusion / Points to Memorize:

- **Amazon S3 Prefixes:** Using prefixes in an S3 bucket allows parallelizing reads and writes, which improves performance and scalability.
- **Performance Scaling:** Amazon S3 can handle thousands of requests per second per prefix, so properly designed prefix structures can prevent throttling.
- **Cost Efficiency:** Optimizing S3 performance with prefixes is more resource-efficient and cost-effective compared to creating multiple buckets or using more expensive storage solutions like Amazon EFS.

Question 35

A leading video streaming service delivers billions of hours of content from Amazon Simple Storage Service (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?

Correct option:

Store the intermediary query results in Amazon S3 Standard storage class

Explanation behind correct option:

- **Amazon S3 Standard:** Amazon S3 Standard offers high durability, availability, and performance object storage for frequently accessed data. Because it delivers low latency and high throughput, S3 Standard is appropriate for a wide variety of use cases, including cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics. As there is no minimum storage duration charge and no retrieval fee, S3 Standard is the most cost-effective storage class for intermediary query results that are heavily referenced by other parts of the analytics pipeline and need to be kept only for 24 hours.

Incorrect options and their explanations:

1. **Store the intermediary query results in Amazon S3 Glacier Instant Retrieval storage class:**
   - **Explanation:** Amazon S3 Glacier Instant Retrieval delivers the fastest access to archive storage with low latency and high throughput similar to S3 Standard and S3 Standard-IA. However, it has a minimum storage duration charge of 90 days, making it not cost-effective for data that needs to be kept only for 24 hours.
2. **Store the intermediary query results in Amazon S3 Standard-Infrequent Access (IA) storage class:**
   - **Explanation:** Amazon S3 Standard-IA is designed for data that is accessed less frequently but requires rapid access when needed. It has a minimum storage duration charge of 30 days and retrieval fees. Given that the intermediary query results need to be kept only for 24 hours and are heavily referenced, S3 Standard-IA is not cost-effective.
3. **Store the intermediary query results in Amazon S3 One Zone-Infrequent Access (IA) storage class:**
   - **Explanation:** Amazon S3 One Zone-IA is designed for data that is accessed less frequently but requires rapid access when needed and stores data in a single availability zone. It also has a minimum storage duration charge of 30 days. For data that needs to be stored only for 24 hours and accessed frequently, this is not cost-effective.

Conclusion / Points to Memorize:

- **S3 Standard:** Best for frequently accessed data with no minimum storage duration charge and no retrieval fee.
- **S3 Glacier Instant Retrieval, S3 Standard-IA, and S3 One Zone-IA:** These options have minimum storage duration charges and retrieval fees, making them less cost-effective for short-term, frequently accessed data.
- **Storage Class Selection:** Always consider the minimum storage duration and retrieval costs when selecting a storage class for frequently accessed, short-term data.

Question 36

A healthcare startup needs to enforce compliance and regulatory guidelines for objects stored in Amazon S3. One of the key requirements is to provide adequate protection against accidental deletion of objects. As a solutions architect, what are your recommendations to address these guidelines? (Select two)

Correct options

1. Enable versioning on the Amazon S3 bucket
2. Enable multi-factor authentication (MFA) delete on the Amazon S3 bucket

Explanation behind correct options:

1. **Enable versioning on the Amazon S3 bucket:**

Versioning is a means of keeping multiple variants of an object in the same bucket. With versioning, you can preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite. If you delete an object, instead of removing it permanently, Amazon S3 inserts a delete marker, which becomes the current object version. You can always restore the previous version.

2. **Enable multi-factor authentication (MFA) delete on the Amazon S3 bucket:**

To provide additional protection, multi-factor authentication (MFA) delete can be enabled. MFA delete requires secondary authentication to take place before objects can be permanently deleted from an Amazon S3 bucket. This adds an extra layer of security to prevent accidental deletions.

Incorrect options and their explanations:

1. **Create an event trigger on deleting any Amazon S3 object. The event invokes an Amazon Simple Notification Service (Amazon SNS) notification via email to the IT manager:**

Sending an event trigger after object deletion does not meet the objective of preventing object deletion by mistake because the object has already been deleted. This option does not prevent accidental deletion but merely notifies after the event.

2. **Establish a process to get managerial approval for deleting Amazon S3 objects:**

This option suggests a manual process that relies on human intervention and does not leverage the automated protection features provided by AWS S3, making it less effective and reliable.

3. **Change the configuration on Amazon S3 console so that the user needs to provide additional confirmation while deleting any Amazon S3 object:**

There is no provision to set up Amazon S3 configuration to ask for additional confirmation before deleting an object. This feature does not exist in Amazon S3.

Conclusion / Points to Memorize:

- **Versioning:** Helps in keeping multiple variants of an object, enabling easy recovery from accidental deletions.
- **MFA Delete:** Adds an extra layer of security by requiring secondary authentication before permanent deletion of objects.
- **Notification and Manual Processes:** While useful for alerts and approvals, they do not prevent accidental deletions and are not substitutes for automated protections like versioning and MFA delete.

Question 37

The IT department at a consulting firm is conducting a training workshop for new developers. As part of an evaluation exercise on Amazon S3, the new developers were asked to identify the invalid storage class lifecycle transitions for objects stored on Amazon S3. Can you spot the INVALID lifecycle transitions from the options below? (Select two)

Correct options:

1. Amazon S3 Intelligent-Tiering => Amazon S3 Standard
2. Amazon S3 One Zone-IA => Amazon S3 Standard-IA

Explanation behind correct options:

Amazon S3 supports specific lifecycle transitions between storage classes. However, certain transitions are not supported:

- **Amazon S3 Intelligent-Tiering to Amazon S3 Standard:** This transition is invalid because once data is in Intelligent-Tiering, it cannot be moved back to the Standard storage class.
- **Amazon S3 One Zone-IA to Amazon S3 Standard-IA:** This transition is invalid because Amazon S3 One Zone-IA does not support transitioning to the Standard-IA storage class.

Incorrect options and their explanations:

1. **Amazon S3 Standard => Amazon S3 Intelligent-Tiering:**

This is a valid transition. You can move objects from the Standard storage class to Intelligent-Tiering.

2. **Amazon S3 Standard-IA => Amazon S3 Intelligent-Tiering:**

This is also a valid transition. Objects can be moved from Standard-IA to Intelligent-Tiering.

3. **Amazon S3 Standard-IA => Amazon S3 One Zone-IA:**

This is a valid transition as well. Objects can be transitioned from Standard-IA to One Zone-IA.

Conclusion / Points to Memorize:

- Valid lifecycle transitions include:
  - S3 Standard to any other storage class.

  - Any storage class to S3 Glacier or S3 Glacier Deep Archive.
  - S3 Standard-IA to S3 Intelligent-Tiering or S3 One Zone-IA.
  - S3 Intelligent-Tiering to S3 One Zone-IA.
  - S3 Glacier to S3 Glacier Deep Archive.
- Invalid lifecycle transitions include:
  - Any storage class to the S3 Standard storage class.
  - Any storage class to the Reduced Redundancy storage class.
  - S3 Intelligent-Tiering to S3 Standard-IA.
  - S3 One Zone-IA to S3 Standard-IA or S3 Intelligent-Tiering.

Question 38

A retail company uses Amazon Elastic Compute Cloud (Amazon EC2) instances, Amazon API Gateway, Amazon RDS, Elastic Load Balancer, and Amazon CloudFront services. To improve the security of these services, the Risk Advisory group has suggested a feasibility check for using the Amazon GuardDuty service. Which of the following would you identify as data sources supported by Amazon GuardDuty?

Correct option:

VPC Flow Logs, Domain Name System (DNS) logs, AWS CloudTrail events

Explanation behind correct option:

Amazon GuardDuty is a threat detection service that continuously monitors for malicious activity and unauthorized behavior to protect your AWS accounts, workloads, and data stored in Amazon S3. It uses machine learning, anomaly detection, and integrated threat intelligence to identify and prioritize potential threats. GuardDuty analyzes events from multiple AWS data sources, such as AWS CloudTrail events, Amazon VPC Flow Logs, and DNS logs.

Incorrect options and their explanations:

1. **VPC Flow Logs, Amazon API Gateway logs, Amazon S3 access logs:**

Amazon GuardDuty does not support Amazon API Gateway logs or Amazon S3 access logs as data sources. It uses VPC Flow Logs, DNS logs, and CloudTrail events.

2. **Elastic Load Balancing logs, Domain Name System (DNS) logs, AWS CloudTrail events:**

Elastic Load Balancing logs are not supported by GuardDuty. The correct data sources are VPC Flow Logs, DNS logs, and CloudTrail events.

3. **Amazon CloudFront logs, Amazon API Gateway logs, AWS CloudTrail events:**

Amazon GuardDuty does not support Amazon CloudFront logs or Amazon API Gateway logs as data sources. It uses VPC Flow Logs, DNS logs, and CloudTrail events.

Conclusion / Points to Memorize:

- Amazon GuardDuty uses VPC Flow Logs, DNS logs, and AWS CloudTrail events as data sources.
- GuardDuty is a threat detection service that continuously monitors for malicious activity and unauthorized behavior.
- It uses machine learning, anomaly detection, and integrated threat intelligence to identify and prioritize potential threats.
- Incorrect data sources such as Amazon API Gateway logs, Amazon S3 access logs, Elastic Load Balancing logs, and Amazon CloudFront logs are not supported by GuardDuty.

Question 39

A company has a web application that runs 24/7 in the production environment. The development team at the company runs a clone of the same application in the dev environment for up to 8 hours every day. The company wants to build the MOST cost-optimal solution by deploying these applications using the best-fit pricing options for Amazon Elastic Compute Cloud (Amazon EC2) instances. What would you recommend?

Correct option:

Use Amazon EC2 reserved instance (RI) for the production application and on-demand instances for the dev application.

Explanation behind correct option:

There are multiple pricing options for EC2 instances, such as On-Demand, Savings Plans, Reserved Instances, and Spot Instances. Amazon EC2 Reserved Instances (RI) provide a significant discount (up to 72%) compared to On-Demand pricing and provide a capacity reservation when used in a specific Availability Zone. For the given use case, you can use Amazon EC2 Reserved Instances for the production application as it runs 24/7, ensuring a 72% discount if you avail a 3-year term. On-demand instances offer the flexibility to only pay for the EC2 instance when it is being used, which is suitable for the dev application that runs up to 8 hours per day.

Incorrect options and their explanations:

1. **Use Amazon EC2 reserved instance (RI) for the production application and spot block instances for the dev application:**

Spot blocks can only be used for up to 6 hours, which does not meet the requirement of the dev application running up to 8 hours. Moreover, AWS has stopped offering Spot blocks to new customers.

2. **Use Amazon EC2 reserved instance (RI) for the production application and spot instances for the dev application:**

Spot instances can be reclaimed by AWS with a two-minute notice, making them unreliable for running the dev application that needs up to 8 hours of uptime.

3. **Use on-demand Amazon EC2 instances for the production application and spot instances for the dev application:**

Using On-Demand instances for the production application would be more expensive compared to Reserved Instances. Spot instances, as mentioned, are unreliable for running the dev application due to the potential for interruption.

Conclusion / Points to Memorize:

- Use Reserved Instances for applications that run continuously (24/7) to benefit from cost savings.
- On-Demand instances are suitable for applications with variable usage patterns and short-term needs.
- Spot instances are cost-effective but not suitable for applications requiring consistent uptime due to potential interruptions.
    - Spot blocks are no longer available for new customers and have a maximum duration of 6 hours.

Question 40

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

Correct Option:

- **Upload the compressed file using multipart upload with Amazon S3 Transfer Acceleration (Amazon S3TA)**

Explanation Behind Correct Option:

- **Upload the compressed file using multipart upload with Amazon S3 Transfer Acceleration (Amazon S3TA):** Amazon S3 Transfer Acceleration (Amazon S3TA) enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket. Transfer Acceleration uses Amazon CloudFront's globally distributed edge locations to route data to Amazon S3 over an optimized network path, significantly speeding up the transfer process. Multipart upload allows a single object to be uploaded as a set of parts, which can be uploaded independently and in parallel, maximizing the use of available bandwidth and improving throughput. This combination of multipart upload and S3 Transfer Acceleration provides the fastest upload method.

Incorrect Options and Their Brief Explanation:

- **Upload the compressed file in a single operation:** Uploading large objects in a single operation can be slow and inefficient, especially when the object size exceeds 100 megabytes. Multipart upload is preferred for large files as it improves throughput by allowing parallel uploads of parts. Therefore, this option is not suitable for a 2-gigabyte file.
- **Upload the compressed file using multipart upload:** While multipart upload is faster than a single operation, using Amazon S3 Transfer Acceleration (Amazon S3TA) in conjunction with multipart upload would further enhance the upload speed, making this option less optimal.
- **FTP the compressed file into an Amazon EC2 instance that runs in the same region as the Amazon S3 bucket. Then transfer the file from the Amazon EC2 instance into the Amazon S3 bucket:** This method involves unnecessary complexity and additional steps, making it less efficient and slower compared to using multipart upload with S3 Transfer Acceleration.

Conclusion / Points to Memorize:

- For large file uploads (greater than 100 MB), use multipart upload to improve throughput and resiliency.
- Amazon S3 Transfer Acceleration can significantly speed up file transfers by using optimized network paths via Amazon CloudFront's edge locations.

• Combining multipart upload with Amazon S3 Transfer Acceleration offers the fastest and most efficient method for uploading large files to Amazon S3.

Question 41

The flagship application for a gaming company connects to an Amazon Aurora database and the entire technology stack is currently deployed in the United States. Now, the company has plans to expand to Europe and Asia for its operations. It needs the games table to be accessible globally but needs the users and games_played tables to be regional only. How would you implement this with minimal application refactoring?

Correct Option:

- **Use an Amazon Aurora Global Database for the games table and use Amazon Aurora for the users and games_played tables**

Explanation Behind Correct Option:

- **Use an Amazon Aurora Global Database for the games table and use Amazon Aurora for the users and games_played tables:** Amazon Aurora Global Database is designed for globally distributed applications, allowing a single Amazon Aurora database to span multiple AWS regions. It replicates data with no impact on database performance, enables fast local reads with low latency in each region, and provides disaster recovery from region-wide outages. This setup allows the games table to be globally accessible while keeping the users and games_played tables regional, meeting the requirement with minimal application refactoring.

Incorrect Options and Their Brief Explanation:

- **Use an Amazon Aurora Global Database for the games table and use Amazon DynamoDB tables for the users and games_played tables:** This option involves using two different database services (Aurora for games table and DynamoDB for users and games_played tables), which would require significant application refactoring due to the different APIs and data models of SQL and NoSQL databases.
- **Use an Amazon DynamoDB global table for the games table and use Amazon Aurora for the users and games_played tables:** Similar to the previous option, this involves mixing SQL and NoSQL databases, which would complicate the application logic and require extensive refactoring.
- **Use an Amazon DynamoDB global table for the games table and use Amazon DynamoDB tables for the users and games_played tables:** This option would require moving the entire data model to DynamoDB, which is a NoSQL database, and would require significant changes to the application's database interaction layer, resulting in high refactoring effort.

Conclusion / Points to Memorize:

- Amazon Aurora Global Database is suitable for globally distributed applications, allowing seamless replication across regions with minimal impact on performance.
- Using the same database service (Amazon Aurora) for both global and regional tables minimizes the need for application refactoring.
- Mixing SQL (Amazon Aurora) and NoSQL (Amazon DynamoDB) databases can significantly increase the complexity and refactoring effort of an application.

Question 42

A company is in the process of migrating its on-premises SMB file shares to AWS so the company can get out of the business of managing multiple file servers across dozens of offices. The company has 200 terabytes of data in its file servers. The existing on-premises applications and native Windows workloads should continue to have low latency access to this data which needs to be stored on a file system service without any disruptions after the migration. The company also wants any new applications deployed on AWS to have access to this migrated data. Which of the following is the best solution to meet this requirement?

Correct Option:

- **Use Amazon FSx File Gateway to provide low-latency, on-premises access to fully managed file shares in Amazon FSx for Windows File Server. The applications deployed on AWS can access this data directly from Amazon FSx in AWS**

Explanation Behind Correct Option:

- **Use Amazon FSx File Gateway to provide low-latency, on-premises access to fully managed file shares in Amazon FSx for Windows File Server. The applications deployed on AWS can access this data directly from Amazon FSx in AWS:** Amazon FSx File Gateway provides low-latency access to fully managed file shares in Amazon FSx for Windows File Server, making it suitable for on-premises applications and native Windows workloads. Applications deployed on AWS can also access these file shares directly from Amazon FSx.

Incorrect Options and Their Brief Explanation:

- **Use Amazon Storage Gateway's File Gateway to provide low-latency, on-premises access to fully managed file shares in Amazon FSx for Windows File Server. The applications deployed on AWS can access this data directly from Amazon FSx in AWS:** AWS Storage Gateway's File Gateway does not support file shares in Amazon FSx for Windows File Server, making this option incorrect.
- **Use AWS Storage Gateway's File Gateway to provide low-latency, on-premises access to fully managed file shares in Amazon S3. The applications deployed on AWS can access this data directly from Amazon S3:** S3 is an object storage service, not a file system service, which does not meet the requirement for low-latency access and file system protocol support.
- **Use Amazon FSx File Gateway to provide low-latency, on-premises access to fully managed file shares in Amazon EFS. The applications deployed on AWS can access this data directly from Amazon EFS:** Amazon FSx File Gateway does not support Amazon EFS, and EFS does not support SMB protocol, making this option incorrect.

Conclusion / Points to Memorize:

- Use Amazon FSx File Gateway for low-latency access to Amazon FSx for Windows File Server.
- Ensure compatibility with SMB protocol for native Windows workloads.
    - Differentiate between object storage services (S3) and file system services (FSx, EFS) based on use case requirements.

Question 43

A major bank is using Amazon Simple Queue Service (Amazon 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?

Correct Option:

Use Amazon SQS FIFO (First-In-First-Out) queue in batch mode of 4 messages per operation to process the messages at the peak rate.

Explanation Behind Correct Option:

- Amazon Simple Queue Service (SQS) offers two types of message queues: Standard queues and FIFO (First-In-First-Out) queues.
- FIFO queues ensure that messages are processed in the exact order they are sent and are not delivered more than once.
- FIFO queues support up to 300 transactions per second (TPS) without batching. However, using batching, they support up to 3,000 TPS, where each batch can contain up to 10 messages.
- To handle 1000 messages per second, using a FIFO queue in batch mode of 4 messages per operation ensures that the FIFO queue can support up to 1200 messages per second (300 TPS * 4 messages).

Incorrect Options and Their Brief Explanation:

1. **Use Amazon SQS standard queue to process the messages:**
    - Incorrect because standard queues do not guarantee the order of message processing, which is a requirement in this use case.
2. **Use Amazon SQS FIFO (First-In-First-Out) queue to process the messages:**
    - Incorrect because by default, FIFO queues support only up to 300 messages per second without batching, which is insufficient for the peak rate of 1000 messages per second.
3. **Use Amazon SQS FIFO (First-In-First-Out) queue in batch mode of 2 messages per operation to process the messages at the peak rate:**
    - Incorrect because batching only 2 messages per operation would support up to 600 messages per second (300 TPS * 2 messages), which is insufficient for the required peak rate.

Conclusion / Points to Memorize:

- FIFO queues ensure strict message order and exactly-once processing.
- Batching messages in FIFO queues can significantly increase the throughput.
- For handling high message rates, determine the optimal batch size to meet the required throughput.
    - Use FIFO queues in batch mode to handle high throughput while maintaining message order.

Question 44:

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)

Correct Options:

1. Different versions of a single object can have different retention modes and periods.
2. When you apply a retention period to an object version explicitly, you specify a Retain Until Date for the object version.

Explanation Behind Correct Options:

1. **Different versions of a single object can have different retention modes and periods:**
    - Retention periods apply to individual object versions. This means different versions of the same object can have different retention modes and periods. For example, an older version of an object might have a 30-day retention period, while a newer version might have a 60-day retention period.
2. **When you apply a retention period to an object version explicitly, you specify a Retain Until Date for the object version:**
    - You can set a retention period on an object version explicitly by specifying a Retain Until Date. Amazon S3 stores this date in the object's metadata, protecting the object version until the specified date is reached.

Incorrect Options and Their Brief Explanation:

1. **You cannot place a retention period on an object version through a bucket default setting:**
    - Incorrect because you can set a retention period on an object version either explicitly or through bucket default settings.
2. **When you use bucket default settings, you specify a Retain Until Date for the object version:**
    - Incorrect because, with bucket default settings, you specify a duration (in days or years) for which every object version in the bucket should be protected, not a specific Retain Until Date.
3. **The bucket default settings will override any explicit retention mode or period you request on an object version:**
    - Incorrect because an explicit retention mode and period specified in a request will override any bucket default settings for that particular object version.

Conclusion / Points to Memorize:

- Retention periods in Amazon S3 can be set explicitly or through bucket default settings.
- Different versions of the same object can have different retention periods.
- Explicit retention settings override bucket default settings.
- Use explicit Retain Until Dates for precise control over object version retention.
    - Ensure understanding of retention periods to avoid unexpected behavior with object retention.

Question 45:

An IT consultant is helping the owner of a medium-sized business set up an AWS account. What are the security recommendations he must follow while creating the AWS account root user? (Select two)

Correct Options:

1. Enable Multi Factor Authentication (MFA) for the AWS account root user account
2. Create a strong password for the AWS account root user

Explanation:

1. **Create a strong password for the AWS account root user:**
    - Using a strong password helps protect account-level access to the AWS Management Console. It is one of the fundamental security practices to ensure that the root account is secure from unauthorized access.
2. **Enable Multi Factor Authentication (MFA) for the AWS account root user account:**
    - Enabling MFA adds an extra layer of security by requiring not only a password and username but also something that only the user has on them, i.e., an MFA device. This significantly reduces the risk of unauthorized access to the AWS account root user.

Incorrect Options:

1. **Create AWS account root user access keys and share those keys only with the business owner:**
    - AWS recommends not creating access keys for the root user unless absolutely necessary, and sharing such keys poses a significant security risk. Root access should be limited to avoid potential misuse.
2. **Send an email to the business owner with details of the login username and password for the AWS root user. This will help the business owner to troubleshoot any login issues in future:**
    - AWS advises never sharing the AWS account root user password or access keys with anyone. Sending login credentials via email is insecure and can lead to unauthorized access if the email is intercepted or accessed by someone else.
3. **Encrypt the access keys and save them on Amazon S3:**

- While encryption is good practice, storing access keys, even if encrypted, is not recommended for the root user. If you don't already have an access key for your AWS account root user, don't create one unless you absolutely need to. Access keys for the root user should be avoided to reduce security risks.

Question 46:

A logistics company is building a multi-tier application to track the location of its trucks during peak operating hours. The company wants these data points to be accessible in real-time in its analytics platform via a REST API. The company has hired you as an AWS Certified Solutions Architect Associate to build a multi-tier solution to store and retrieve this location data for analysis. Which of the following options addresses the given use case?

Options:

1. Leverage Amazon API Gateway with AWS Lambda

2. Leverage Amazon API Gateway with Amazon Kinesis Data Analytics (Correct)
3. Leverage Amazon QuickSight with Amazon Redshift
4. Leverage Amazon Athena with Amazon S3

Correct Option:

Leverage Amazon API Gateway with Amazon Kinesis Data Analytics

Explanation:

- **Amazon API Gateway**: A fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. API Gateway can handle thousands of concurrent API calls, enabling real-time communication and data collection.
- **Amazon Kinesis Data Analytics**: A service that allows real-time processing of streaming data. It can be used to transform and analyze the location data streaming from the trucks. Kinesis Data Analytics can handle high throughput and provides low-latency access to streaming data.

For the given use case, Amazon API Gateway can be used to create a REST API that captures location data from trucks. This data can then be sent to Amazon Kinesis Data Analytics for real-time processing and analysis.

Incorrect Options:

- **Leverage Amazon Athena with Amazon S3**: Athena is used for querying data stored in S3 using SQL. It is not suitable for real-time data streaming or processing, and cannot create a REST API for real-time data ingestion.
- **Leverage Amazon QuickSight with Amazon Redshift**: QuickSight is a business intelligence service for data visualization and analysis, and Redshift is a data warehouse service. These services are not designed for real-time data ingestion and processing through a REST API.
- **Leverage Amazon API Gateway with AWS Lambda**: While Lambda can process data, it is not optimized for continuous real-time data streams and large-scale analytics compared to Kinesis Data Analytics. Lambda is more suited for discrete, event-driven applications rather than continuous data streams.

Question 47:

A gaming company is developing a mobile game that streams score updates to a backend processor and then publishes results on a leaderboard. The company has hired you as an AWS Certified Solutions Architect Associate to design a solution that can handle major traffic spikes, process the mobile game updates in the order of receipt, and store the processed updates in a highly available database. The company wants to minimize the management overhead required to maintain the solution.

Correct Answer:

Push score updates to Amazon Kinesis Data Streams, which uses an AWS Lambda function to process these updates and then store these processed updates in Amazon DynamoDB.

Explanation and Context:

- **Amazon Kinesis Data Streams (KDS):** KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources. It ensures the ordering of records and allows multiple Amazon Kinesis applications to read and replay records in the same order.
- **AWS Lambda Integration:** AWS Lambda natively integrates with KDS, handling polling, checkpointing, and error handling complexities. This reduces management overhead since there is no need to manage the underlying infrastructure.
- **Amazon DynamoDB:** A highly available, fully managed NoSQL database that scales seamlessly, making it ideal for storing processed updates.

Incorrect Options and Reasons:

1. **Push score updates to an Amazon Simple Queue Service (Amazon SQS) queue which uses a fleet of Amazon EC2 instances (with Auto Scaling) to process these updates in the Amazon SQS queue and then store these processed updates in an Amazon RDS MySQL database:**
   - Involves managing Amazon EC2 instances and Auto Scaling, which adds management overhead.
   - Amazon RDS MySQL database may not handle the same scale and availability as DynamoDB for this use case.
2. **Push score updates to Amazon Kinesis Data Streams which uses a fleet of Amazon EC2 instances (with Auto Scaling) to process the updates in Amazon Kinesis Data Streams and then store these processed updates in Amazon DynamoDB:**
   - Using Amazon EC2 instances for processing adds significant management overhead, which contradicts the requirement to minimize it.
3. **Push score updates to an Amazon Simple Notification Service (Amazon SNS) topic, subscribe an AWS Lambda function to this Amazon SNS topic to process the updates and then store these processed updates in a SQL database running on Amazon EC2 instance:**
   - Involves managing Amazon EC2 instances, which increases management overhead.
   - SQL databases may not handle the high throughput and availability requirements as efficiently as DynamoDB for this use case.

Conclusion or Points to Memorize:

- **Amazon Kinesis Data Streams** is suitable for real-time data ingestion and ensures ordered processing of records.
- **AWS Lambda** integration with KDS abstracts away the infrastructure management, reducing the overall management overhead.

- **Amazon DynamoDB** provides a highly available and scalable database solution, ideal for storing processed updates from Lambda.

Question 48:

A media agency stores its re-creatable assets on Amazon Simple Storage Service (Amazon S3) buckets. The assets are accessed by a large number of users for the first few days, and the frequency of access falls drastically after a week. Although the assets would be accessed occasionally after the first week, they must continue to be immediately accessible when required. The cost of maintaining all the assets on Amazon S3 storage is turning out to be very expensive, and the agency is looking at reducing costs as much as possible.

Correct Answer:

Configure a lifecycle policy to transition the objects to Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA) after 30 days.

Explanation and Context:

- **Amazon S3 One Zone-IA:** This storage class is for data that is accessed less frequently but requires rapid access when needed. Unlike other S3 storage classes, which store data in a minimum of three Availability Zones (AZs), S3 One Zone-IA stores data in a single AZ, reducing costs by 20% compared to S3 Standard-IA. It is ideal for infrequently accessed and re-creatable data where availability and resilience across multiple AZs are not critical.
- **Minimum Storage Duration:** The minimum storage duration for transitioning objects from S3 Standard to S3 One Zone-IA is 30 days. This ensures that frequently accessed data remains in S3 Standard for the initial high-access period before moving to a lower-cost storage class.

Incorrect Options and Reasons:

1. **Configure a lifecycle policy to transition the objects to Amazon S3 Standard-Infrequent Access (S3 Standard-IA) after 7 days:**
   - The minimum storage duration for transitioning objects to S3 Standard-IA is 30 days, making this option incorrect.
2. **Configure a lifecycle policy to transition the objects to Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA) after 7 days:**
   - Similar to S3 Standard-IA, the minimum storage duration for transitioning objects to S3 One Zone-IA is 30 days, so this option is incorrect.
3. **Configure a lifecycle policy to transition the objects to Amazon S3 Standard-Infrequent Access (S3 Standard-IA) after 30 days:**
   - Although S3 Standard-IA is suitable for infrequently accessed data and provides high durability and low latency, it costs more than S3 One Zone-IA due to the redundant storage across multiple AZs. Given that the data is re-creatable, the additional cost for multi-AZ redundancy is unnecessary.

Conclusion or Points to Memorize:

- **Amazon S3 One Zone-IA** is the most cost-effective solution for infrequently accessed, re-creatable data that must remain rapidly accessible when needed.
- Ensure that data remains in S3 Standard for at least 30 days before transitioning to a lower-cost storage class like S3 One Zone-IA to meet the minimum storage duration requirements.

Question 49:

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 Amazon Elastic File System (Amazon EFS) created in the 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.

Correct Answer:

The spreadsheet on the Amazon Elastic File System (Amazon EFS) can be accessed in other AWS regions by using an inter-region VPC peering connection.

Explanation and Context:

- **Amazon Elastic File System (Amazon EFS):** EFS provides a simple, scalable, fully managed elastic NFS file system for use with AWS Cloud services and on-premises resources. It is a regional service that stores data within and across multiple Availability Zones (AZs) for high availability and durability.
- **Inter-region VPC Peering:** By establishing an inter-region VPC peering connection, EC2 instances in different regions can access the EFS file system. This approach allows the sourcing team in various regions to collaborate on the spreadsheet with minimal operational overhead.

Incorrect Options and Reasons:

1. **The spreadsheet will have to be copied in Amazon S3 which can then be accessed from any AWS region:**
   - **Amazon S3** does not support in-place edits and is not POSIX compliant. This would require developing a custom application to simulate in-place edits, which adds significant operational overhead.
2. **The spreadsheet data will have to be moved into an Amazon RDS for MySQL database which can then be accessed from any AWS region:**

- **Amazon RDS** is a relational database service. Moving the spreadsheet data into RDS would require custom code to replicate spreadsheet functionality, adding considerable complexity and operational overhead.

3. **The spreadsheet will have to be copied into Amazon EFS file systems of other AWS regions as Amazon EFS is a regional service and it does not allow access from other AWS regions:**
   - Copying the spreadsheet into separate EFS file systems in each region would prevent collaboration, as each team would work on its own file instead of a single, unified file. This defeats the purpose of enabling collaborative editing.

Conclusion or Points to Memorize:

- **Inter-region VPC Peering** with Amazon EFS is an effective solution to enable multi-region collaboration on a single file, leveraging the scalability and managed nature of EFS.

- Choosing solutions that minimize operational overhead and maintain the integrity of collaborative workflows is essential for efficient and effective cloud architecture.

Question 50:

A company uses Amazon DynamoDB as a data store for various kinds of customer data, such as user profiles, user events, clicks, and visited links. Some of these use-cases require a high request rate (millions of requests per second), low predictable latency, and reliability. The company now wants to add a caching layer to support high read volumes.

Correct Answers:

1. Amazon DynamoDB Accelerator (DAX)
2. Amazon ElastiCache

Explanation and Context:

1. **Amazon DynamoDB Accelerator (DAX):**
   - **DAX** is a fully managed, highly available, in-memory cache for DynamoDB. It provides up to a 10x performance improvement, reducing read response times from milliseconds to microseconds, even at millions of requests per second. DAX is specifically designed to seamlessly integrate with DynamoDB, offering a significant boost in performance without requiring additional code changes for cache management.
2. **Amazon ElastiCache:**
   - **ElastiCache** supports both Redis and Memcached and acts as an in-memory cache to reduce the load on your databases. For applications with extremely high request rates and/or low latency requirements, ElastiCache provides a high-performance middle tier. It is an ideal caching layer for DynamoDB, offering scalable, managed caching solutions to improve read throughput and reduce latency.

Incorrect Options and Reasons:

1. **Amazon Relational Database Service (Amazon RDS):**
   - **RDS** is a managed relational database service. It is not designed to act as a caching layer for DynamoDB and does not provide the in-memory caching benefits required for this use case.
2. **Amazon OpenSearch Service:**
   - **OpenSearch Service** is a managed service for search and analytics workloads derived from Elasticsearch. It is not suitable for use as a caching layer for DynamoDB due to its primary focus on search and analytics, rather than caching.
3. **Amazon Redshift:**
   - **Redshift** is a managed data warehouse service designed for large-scale data analysis. It is not intended to function as a caching layer for DynamoDB, making it an unsuitable choice for this scenario.

Conclusion or Points to Memorize:

- **Amazon DynamoDB Accelerator (DAX)** and **Amazon ElastiCache** are the optimal solutions for implementing a high-performance caching layer for DynamoDB.

- These services enhance read performance by providing low latency, high throughput, and seamless integration with existing DynamoDB workloads.

Question 51:

A new DevOps engineer has joined a large financial services company recently. As part of his onboarding, the IT department is conducting a review of the checklist for tasks related to AWS Identity and Access Management (AWS IAM).

Correct Answers:

1. Enable AWS Multi-Factor Authentication (AWS MFA) for privileged users

2. Configure AWS CloudTrail to log all AWS Identity and Access Management (AWS IAM) actions

Explanation and Context:

1. **Enable AWS Multi-Factor Authentication (AWS MFA) for privileged users:**
   - **MFA adds an extra layer of security** by requiring not just a username and password but also something that the user physically possesses (like an MFA-enabled mobile device or hardware MFA token). This reduces the risk of compromised credentials.
2. **Configure AWS CloudTrail to log all AWS Identity and Access Management (AWS IAM) actions:**
   - **CloudTrail logging is essential** for security monitoring and auditing. By enabling CloudTrail, you can track all IAM activities, helping to identify any unauthorized access attempts or other suspicious activities.

Incorrect Options and Reasons:

1. **Grant maximum privileges to avoid assigning privileges again:**
   - **AWS recommends the principle of least privilege**, which means granting only the permissions necessary to perform a task. Granting maximum privileges increases the risk of misuse and security breaches.
2. **Create a minimum number of accounts and share these account credentials among employees:**
   - **Sharing account credentials** is against AWS best practices as it compromises security. Each user should have their own individual account for accountability and security.
3. **Use user credentials to provide access specific permissions for Amazon EC2 instances:**
   - **Using IAM roles** to grant access permissions to EC2 instances is recommended instead of user credentials. This approach provides temporary security credentials and avoids the need for hardcoding sensitive credentials.

Conclusion or Points to Memorize:

- **Enabling MFA** and **logging IAM actions** are critical security best practices to ensure robust access management and monitoring within AWS environments.
- Following **least privilege principles** and using **IAM roles** for resource access are key to maintaining a secure and efficient IAM setup.

Question 52:

The engineering team at an e-commerce company wants to establish a dedicated, encrypted, low latency, and high throughput connection between its data center and AWS Cloud. The engineering team has set aside sufficient time to account for the operational overhead of establishing this connection.

Correct Answer:

Use AWS Direct Connect plus virtual private network (VPN) to establish a connection between the data center and AWS Cloud.

Explanation and Context:

- **AWS Direct Connect + VPN:** This combination allows the company to establish a dedicated network connection with the low latency and high throughput benefits of AWS Direct Connect, along with the encryption benefits of a VPN. AWS Direct Connect establishes a private, dedicated connection between the data center and AWS. Adding a VPN ensures that the data transmitted over this connection is encrypted using IPsec, providing an additional layer of security.

Incorrect Options and Reasons:

1. **Use AWS site-to-site VPN to establish a connection between the data center and AWS Cloud:** Site-to-site VPN provides secure connectivity but is dependent on the public internet, which can result in higher latency and less consistent throughput compared to Direct Connect.
2. **Use AWS Transit Gateway to establish a connection between the data center and AWS Cloud:** AWS Transit Gateway is used for interconnecting VPCs and on-premises networks but does not itself establish a low latency, high throughput connection directly between a data center and AWS.
3. **Use AWS Direct Connect to establish a connection between the data center and AWS Cloud:** While AWS Direct Connect provides low latency and high throughput, it does not provide encryption by itself. The addition of a VPN ensures data security over the connection.

Conclusion or Points to Memorize:

- **AWS Direct Connect + VPN** offers the best combination of low latency, high throughput, and encrypted connections, suitable for secure and high-performance connectivity between on-premises data centers and AWS Cloud.
- Understanding the specific benefits and limitations of AWS networking services is crucial for designing solutions that meet both performance and security requirements.

Question 53:

A large financial institution operates an on-premises data center with hundreds of petabytes of data managed on Microsoft's Distributed File System (DFS). The CTO wants the organization to transition into a hybrid cloud environment and run data-intensive analytics workloads that support DFS.

Correct Answer:

Amazon FSx for Windows File Server

Explanation and Context:

- **Amazon FSx for Windows File Server:** This service provides fully managed, highly reliable file storage accessible over the SMB protocol. It integrates with Microsoft Active Directory and supports DFS, making it ideal for transitioning existing on-premises DFS workloads to the cloud. It scales to handle hundreds of petabytes of data and is built on Windows Server, ensuring compatibility with existing Windows-based applications and tools.

Incorrect Options and Reasons:

1. **Amazon FSx for Lustre:** While FSx for Lustre is designed for high-performance workloads such as HPC and machine learning, it does not support Microsoft DFS, making it unsuitable for this use case.
2. **AWS Directory Service for Microsoft Active Directory (AWS Managed Microsoft AD):** This service provides managed Active Directory in the AWS cloud but does not support DFS, thus it cannot facilitate the migration of DFS-based workloads.
3. **Microsoft SQL Server on AWS:** This service allows running SQL Server databases on AWS but does not support DFS, making it irrelevant for this scenario.

Conclusion or Points to Memorize:

- **Amazon FSx for Windows File Server** is the optimal choice for migrating and managing Microsoft DFS-based workloads in a hybrid cloud environment.
  - It integrates seamlessly with Active Directory and supports the SMB protocol, ensuring compatibility with existing Windows-based infrastructures.

Question 54:

An audit department generates and accesses the audit reports only twice in a financial year. The department uses AWS Step Functions to orchestrate the report creating process that has failover and retry scenarios built into the solution. The underlying data to create these audit reports is stored on Amazon S3, runs into hundreds of Terabytes and should be available with millisecond latency.

Correct Answer:

Amazon S3 Standard-Infrequent Access (S3 Standard-IA)

Explanation and Context:

- **Amazon S3 Standard-IA:** This storage class is designed for data that is infrequently accessed but needs to be retrieved rapidly when required. It offers a low per GB storage price and a retrieval fee, making it cost-effective for data that is accessed occasionally. It provides millisecond latency and high durability (99.999999999%), with slightly lower availability (99.9%) compared to S3 Standard.

Incorrect Options and Reasons:

1. **Amazon S3 Standard:** Designed for frequently accessed data with higher availability (99.99%) but is more expensive for infrequently accessed data.
2. **Amazon S3 Intelligent-Tiering:** Optimizes costs by automatically moving data between access tiers based on usage patterns. While efficient, S3 Standard-IA is more cost-effective for known infrequent access patterns.
3. **Amazon S3 Glacier Deep Archive:** Offers low-cost storage for long-term archiving but does not provide millisecond latency, making it unsuitable for this use case.

Conclusion or Points to Memorize:

- **S3 Standard-IA** is ideal for infrequent access scenarios where data needs to be retrieved quickly when accessed.
  - **Cost-effectiveness:** Lower storage costs combined with retrieval fees make it suitable for large datasets accessed a few times a year.

Question 55:

An IT security consultancy is working on a solution to protect data stored in Amazon S3 from any malicious activity as well as check for any vulnerabilities on Amazon EC2 instances.

Correct Answer:

Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3. Use security assessments provided by Amazon Inspector to check for vulnerabilities on Amazon EC2 instances.

Explanation and Context:

- **Amazon GuardDuty:** A threat detection service that continuously monitors for malicious or unauthorized behavior to help protect your AWS accounts and data stored in Amazon S3. It analyzes data from AWS CloudTrail, VPC Flow Logs, and DNS logs using machine learning and threat intelligence to identify and prioritize potential threats.
- **Amazon Inspector:** An automated security assessment service that helps improve the security and compliance of applications deployed on Amazon EC2 by checking for vulnerabilities and unintended network accessibility. It provides detailed reports with findings prioritized by level of severity.

Incorrect Options and Reasons:

1. **Use Amazon Inspector to monitor any malicious activity on data stored in Amazon S3. Use security assessments provided by Amazon Inspector to check for vulnerabilities on Amazon EC2 instances:** Amazon Inspector does not monitor malicious activity on S3. It is designed to assess security vulnerabilities on EC2 instances.
2. **Use Amazon Inspector to monitor any malicious activity on data stored in Amazon S3. Use security assessments provided by Amazon GuardDuty to check for vulnerabilities on Amazon EC2 instances:** This option is incorrect because GuardDuty is used for monitoring threats and malicious activity, not for conducting security assessments or checking vulnerabilities.
3. **Use Amazon GuardDuty to monitor any malicious activity on data stored in Amazon S3. Use security assessments provided by Amazon GuardDuty to check for vulnerabilities on Amazon EC2 instances:** GuardDuty does not perform security assessments or check for vulnerabilities on EC2 instances; this is the role of Amazon Inspector.

Conclusion or Points to Memorize:

- Amazon GuardDuty is best suited for detecting and monitoring threats and malicious activities, especially in S3, using intelligent threat detection mechanisms.
- Amazon Inspector provides detailed security assessments and checks for vulnerabilities in EC2 instances, making it essential for maintaining secure cloud infrastructure.

Question 56:

The engineering team at a Spanish professional football club has built a notification system for its website using Amazon Simple Notification Service (Amazon SNS) notifications which are then handled by an AWS 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.

Correct Answer:

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

Explanation and Context:

- **AWS Lambda Concurrency Limits:** AWS Lambda has default concurrency limits (1,000 concurrent executions per AWS account per region), which can be reached during peak periods, such as a football season for the club. When these limits are exceeded, additional Lambda function invocations are throttled, causing messages to not be delivered.
- **Solution:** Contact AWS Support to request an increase in the concurrency quota for Lambda functions to ensure that all notifications are processed without delay during high-demand periods.

Incorrect Options and Reasons:

1. **Amazon SNS has hit a scalability limit, so the team needs to contact AWS support to raise the account limit:** This is incorrect because Amazon SNS is designed to automatically scale with demand and does not have such scalability limits that users need to manually adjust.
2. **The engineering team needs to provision more servers running the Amazon SNS service:** Amazon SNS is a fully managed service that handles scaling automatically; users do not need to provision servers for SNS.
3. **The engineering team needs to provision more servers running the AWS Lambda service:** AWS Lambda is also a serverless, managed service. Users cannot provision "servers" for Lambda as its scaling and server management are handled by AWS.

Conclusion or Points to Memorize:

- Understand the scalability features and limits of AWS services like AWS Lambda, particularly in scenarios with highly variable load.
- In serverless architectures, issues related to scaling often pertain to configurations like concurrency limits, not the provisioning of physical servers.

- Always consider contacting AWS support for limit increases well in advance of anticipated high-load events to ensure service continuity.

Question 57:

A US-based healthcare startup is building an interactive diagnostic tool for COVID-19 related assessments. The users would be required to capture their personal health records via this tool. As this is sensitive health information, the backup of the user data must be kept encrypted in Amazon Simple Storage Service (Amazon S3). The startup does not want to provide its own encryption keys but still wants to maintain an audit trail of when an encryption key was used and by whom.

Correct Answer:

Use server-side encryption with AWS Key Management Service keys (SSE-KMS) to encrypt the user data on Amazon S3.

Explanation and Context:

- **AWS Key Management Service (AWS KMS):** This service provides secure, highly available key management, essential for handling sensitive health information. SSE-KMS not only helps in encrypting the data at rest but also offers an audit trail that records the use of encryption keys, showing when and by whom the keys were used.

Incorrect Options and Reasons:

1. **Use server-side encryption with Amazon S3 managed keys (SSE-S3):** While this option also encrypts data at rest, it does not provide the audit capabilities needed to track key usage, which is crucial for maintaining compliance with health data protection standards.
2. **Use server-side encryption with customer-provided keys (SSE-C):** This method would require the startup to manage its own encryption keys, contrary to their requirement of not wanting to handle key management directly.
3. **Use client-side encryption with client provided keys and then upload the encrypted user data to Amazon S3:** This option is incorrect as the startup specifically indicated a preference for not managing encryption keys directly, focusing instead on server-side solutions.

Conclusion or Points to Memorize:

- For businesses handling sensitive information, such as health records, and requiring compliance with audit trails without the burden of key management, AWS KMS integrated with S3 (SSE-KMS) provides a robust solution.
- Understanding the different encryption options available in S3 and their respective capabilities, especially regarding key management and auditing, is critical for making informed decisions that align with security and compliance requirements.

Question 58:

A leading carmaker would like to build a new car-as-a-sensor service by leveraging fully serverless components that are provisioned and managed automatically by AWS. The development team at the carmaker does not want an option that requires the capacity to be manually provisioned, as it does not want to respond manually to changing volumes of sensor data.

Correct Answer:

Ingest the sensor data in an Amazon Simple Queue Service (Amazon SQS) standard queue, which is polled by an AWS Lambda function in batches and the data is written into an auto-scaled Amazon DynamoDB table for downstream processing.

Explanation and Context:

- **Amazon SQS and AWS Lambda:** This combination is ideal for serverless architectures. SQS can handle incoming sensor data, acting as a buffer and managing the inflow of data without any provisioning requirements. AWS Lambda can automatically process data from SQS as it arrives, scaling automatically with the volume of messages without any manual intervention. This setup allows for efficient handling of variable workloads typical in IoT and sensor-based applications.
- **Auto-scaled Amazon DynamoDB:** As the processed data needs a storage solution, DynamoDB is appropriate for handling large volumes of data with its auto-scaling capabilities, which adjust its throughput capacity based on load, ensuring performance and cost-efficiency.

Incorrect Options and Reasons:

1. **Ingest the sensor data in Amazon Kinesis Data Firehose, which directly writes the data into an auto-scaled Amazon DynamoDB table for downstream processing:** This option is incorrect because Amazon Kinesis Data Firehose does not support direct data writes into DynamoDB. Firehose is designed to stream data to destinations like Amazon S3, Amazon Redshift, or Amazon OpenSearch Service, not DynamoDB.
2. **Ingest the sensor data in an Amazon Simple Queue Service (Amazon SQS) standard queue, which is polled by an application running on an Amazon EC2 instance and the data is written into an auto-scaled Amazon DynamoDB table for downstream processing:** This option

contradicts the requirement for a fully serverless solution as it involves managing EC2 instances, which require manual scaling and capacity management.

3. **Ingest the sensor data in Amazon Kinesis Data Streams, which is polled by an application running on an Amazon EC2 instance and the data is written into an auto-scaled Amazon DynamoDB table for downstream processing:** Similarly, this option involves EC2 instances, which does not align with the carmaker's preference for a serverless architecture that does not require manual provisioning or capacity management.

Conclusion or Points to Memorize:

- The combination of SQS for data ingestion, AWS Lambda for processing, and DynamoDB for storage represents a fully serverless architecture, ideal for applications with variable loads and the need for high scalability without manual infrastructure management.

• Understanding the capabilities and integration possibilities of AWS serverless services is crucial for designing efficient, scalable, and cost-effective solutions for dynamic and data-intensive applications.

Question 59:

A leading social media analytics company is contemplating moving its dockerized application stack into AWS Cloud. The company is not sure about the pricing for using Amazon Elastic Container Service (Amazon ECS) with the EC2 launch type compared to the Amazon Elastic Container Service (Amazon ECS) with the Fargate launch type.

Correct Answer:

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

Explanation and Context:

- **Amazon ECS with EC2 Launch Type:** When using the EC2 launch type, the costs are based on the EC2 instances and any associated EBS volumes you utilize for your application. This means you're responsible for managing the instances, scaling, and their corresponding costs.
- **Amazon ECS with Fargate Launch Type:** Fargate abstracts the underlying server infrastructure, allowing users to focus on deploying and scaling applications without worrying about the underlying hardware. Pricing for Fargate is based solely on the vCPU and memory that your containerized applications are configured to use, not on the underlying instances or storage.

Incorrect Options and Reasons:

1. **Both Amazon ECS with EC2 launch type and Amazon ECS with Fargate launch type are charged based on vCPU and memory resources that the containerized application requests:** This statement is incorrect because the EC2 launch type charges are based on the EC2 instances and EBS volumes used, not directly on vCPU and memory like Fargate.
2. **Both Amazon ECS with EC2 launch type and Amazon ECS with Fargate launch type are just charged based on Elastic Container Service used per hour:** This option is incorrect because it does not accurately represent how ECS charges work. ECS itself does not have a separate hourly charge; instead, charges are based on the resources used (instances, vCPU, memory).
3. **Both Amazon ECS with EC2 launch type and Amazon ECS with Fargate launch type are charged based on Amazon EC2 instances and Amazon EBS Elastic Volumes used:** This is incorrect because, under the Fargate launch type, users do not pay for EC2 instances or EBS volumes; they pay for the compute and memory resources their containerized applications use.

Conclusion or Points to Memorize:

- Understanding the different pricing models of ECS EC2 and ECS Fargate is crucial for cost management when running containerized applications in AWS.

• ECS EC2 involves managing and paying for the instances and volumes, while ECS Fargate offers a simpler pricing model based on the resources the application consumes directly, making it suitable for applications where direct control over servers is less critical.

Question 60:

A healthcare company uses its on-premises infrastructure to run legacy applications that require specialized customizations to the underlying Oracle database as well as its host operating system (OS). The company also wants to improve the availability of the Oracle database layer. The company has hired you as an AWS Certified Solutions Architect – Associate to build a solution on AWS that meets these requirements while minimizing the underlying infrastructure maintenance effort.

Correct Answer:

Leverage multi-AZ configuration of Amazon RDS Custom for Oracle that allows the Database Administrator (DBA) to access and customize the database environment and the underlying operating system.

Explanation and Context:

Amazon RDS Custom for Oracle offers a managed relational database service that also permits customization of the database environment and the OS, which is typically required for legacy applications needing specific patches or configurations. By deploying in a multi-AZ configuration, the solution ensures high availability and fault tolerance, automatically replicating databases across multiple, physically separate AZs and handling failover automatically.

- **High Availability:** Multi-AZ deployments for RDS Custom enhance availability and data durability, crucial for healthcare applications that require constant uptime.
- **Customization:** Unlike standard RDS, RDS Custom allows modifications at the OS and database level, which is necessary for legacy systems that often require specific tweaks to operate effectively.

Incorrect Options and Reasons:

1. **Leverage multi-AZ configuration of Amazon RDS for Oracle that allows the Database Administrator (DBA) to access and customize the database environment and the underlying operating system:** This option is incorrect because standard RDS for Oracle does not allow OS-level or physical server-level customizations.
2. **Leverage cross AZ read-replica configuration of Amazon RDS for Oracle that allows the Database Administrator (DBA) to access and customize the database environment and the underlying operating system:** Cross-AZ read-replicas provide data replication and scalability but do not permit the necessary level of customization of the database environment or OS.
3. **Deploy the Oracle database layer on multiple Amazon EC2 instances spread across two Availability Zones (AZs). This deployment configuration guarantees high availability and also allows the Database Administrator (DBA) to access and customize the database environment and the underlying operating system:** While this option offers high availability and complete control over the database and OS, it requires significant infrastructure management effort, which contradicts the requirement to minimize maintenance efforts.

Conclusion or Points to Memorize:

- Amazon RDS Custom for Oracle is specifically designed to provide the flexibility needed for complex, customized database setups while still offering the benefits of a managed service.

• Choosing RDS Custom for Oracle in a multi-AZ configuration is key for businesses that need a blend of high availability, system resilience, and customization capabilities for legacy applications.

Question 61:

While consolidating logs for the weekly reporting, a development team at an e-commerce company noticed that an unusually large number of illegal AWS application programming interface (API) queries were made sometime during the week. Due to the off-season, there was no visible impact on the systems. However, this event led the management team to seek an automated solution that can trigger near-real-time warnings in case such an event recurs.

Correct Answer:

Create an Amazon CloudWatch metric filter that processes AWS CloudTrail logs having API call details and looks at any errors by factoring in all the error codes that need to be tracked. Create an alarm based on this metric's rate to send an Amazon SNS notification to the required team.

Explanation and Context:

AWS CloudTrail captures a log of all API calls made within an AWS account, providing detailed information about the API calls, including the identity of the API caller, the time of the API call, the source IP address of the API caller, and more. These logs can be processed using Amazon CloudWatch, which can create metric filters to scan for specific error codes or patterns that indicate unauthorized or illegal API usage.

The process involves:

1. **CloudTrail Logs**: These logs are sent to CloudWatch.
2. **Metric Filters**: These are set up in CloudWatch to sift through CloudTrail logs for specific error codes or other signs of unusual or unauthorized activity.
3. **CloudWatch Alarms**: These alarms are triggered based on the criteria set in the metric filters, such as a certain rate of detected errors.

Incorrect Options and Reasons:

1. **Configure AWS CloudTrail to stream event data to Amazon Kinesis. Use Amazon Kinesis stream-level metrics in the Amazon CloudWatch to trigger an AWS Lambda function that will trigger an error workflow:** This is incorrect because AWS CloudTrail does not support direct streaming to Amazon Kinesis. The correct integration path is via CloudWatch or S3.
2. **Run Amazon Athena SQL queries against AWS CloudTrail log files stored in Amazon S3 buckets. Use Amazon QuickSight to generate reports for managerial dashboards:** While this method can provide insights and visualizations for managerial review, it does not offer the near-real-time monitoring and alerting required by the scenario.

3. **AWS Trusted Advisor publishes metrics about check results to Amazon CloudWatch. Create an alarm to track status changes for checks in the Service Limits category for the APIs:** This option focuses on service limits and does not address the detection of unusual API call patterns or errors, which are essential for identifying potential security incidents as described.

Conclusion or Points to Memorize:

- CloudWatch, integrated with CloudTrail, provides a robust mechanism for monitoring and alerting on AWS account activities, capable of detecting and responding to potential security incidents in near real-time.
- Metric filters in CloudWatch enable detailed scrutiny of log data, allowing for the detection of specific events or anomalies that could indicate security issues.

• Amazon SNS can be used in conjunction with CloudWatch alarms to ensure that notifications are promptly sent to the relevant personnel or systems, facilitating rapid response to potential threats.

Question 62:

A retail company's dynamic website is hosted using on-premises servers in the United States. The company is launching its website in Asia, and it wants to optimize the website loading times for new users in Asia. The website's backend must remain in the United States. The website is being launched in a few days, and an immediate solution is needed.

Correct Answer:

Use Amazon CloudFront with a custom origin pointing to the on-premises servers.

Explanation and Context:

Amazon CloudFront is a content delivery network (CDN) that speeds up the distribution of static and dynamic web content to users globally. By setting up CloudFront with a custom origin that points to the on-premises servers, the CDN caches the content at edge locations closer to the users in Asia. This reduces the latency by serving content from a location geographically closer to the users, even though the main servers are located in the United States.

Incorrect Options and Reasons:

1. **Use Amazon CloudFront with a custom origin pointing to the DNS record of the website on Amazon Route 53:** This is incorrect because CloudFront does not use DNS records as origins. It requires direct links to the original servers (custom origins) or services like Amazon S3 for static content.
2. **Migrate the website to Amazon S3. Use S3 cross-region replication (S3 CRR) between AWS Regions in the US and Asia:** This option is not feasible for dynamic websites as Amazon S3 is typically used for static content hosting. S3 does not support server-side processing required by dynamic websites.
3. **Leverage a Amazon Route 53 geo-proximity routing policy pointing to on-premises servers:** This approach does not reduce latency effectively since it still routes traffic back to the servers in the United States, merely optimizing the routing path without caching content closer to the users.

Conclusion or Points to Memorize:

- CloudFront effectively reduces latency for global audiences by caching content at edge locations near the users, making it ideal for quick deployments where backend architecture cannot be altered significantly.

• Understanding the capabilities of CDN technologies like CloudFront and their integration with existing on-premises infrastructure is crucial for optimizing web performance across different geographical locations.

Question 63:

A retail company has developed a REST API which is deployed in an Auto Scaling group behind an Application Load Balancer. The REST API stores the user data in Amazon DynamoDB and any static content, such as images, are served via Amazon Simple Storage Service (Amazon S3). On analyzing the usage trends, it is found that 90% of the read requests are for commonly accessed data across all users.

Correct Answer:

Enable Amazon DynamoDB Accelerator (DAX) for Amazon DynamoDB and Amazon CloudFront for Amazon S3.

Explanation and Context:

- **Amazon DynamoDB Accelerator (DAX):** DAX is a fully managed, highly available, in-memory cache for DynamoDB that provides up to 10x performance improvement—from milliseconds to microseconds—even at millions of requests per second. It is tightly integrated with DynamoDB and is API-compatible, meaning no changes to application code are needed for implementation if the application already uses DynamoDB. This integration makes it an ideal solution for frequently accessed read data, which fits the scenario described where 90% of the reads are for commonly accessed data.

- **Amazon CloudFront:** As a CDN, CloudFront efficiently serves static content stored in Amazon S3. By caching content at edge locations closer to the end-users, CloudFront reduces latency and improves the speed of content delivery, complementing the dynamic data performance improvement provided by DAX.

Incorrect Options and Reasons:

1. **Enable ElastiCache Redis for DynamoDB and Amazon CloudFront for Amazon S3:** While Redis is a powerful in-memory data store that supports complex data types and atomic operations, it is not natively integrated with DynamoDB like DAX. Implementing Redis would require additional management and configuration, which is less efficient compared to the seamless integration offered by DAX.
2. **Enable Amazon DynamoDB Accelerator (DAX) for Amazon DynamoDB and ElastiCache Memcached for Amazon S3:** ElastiCache Memcached is typically used for caching database queries and object caching; however, it is not suitable for caching static content from S3, which is better served by a CDN like CloudFront.
3. **Enable ElastiCache Redis for DynamoDB and ElastiCache Memcached for Amazon S3:** Similar to the first incorrect option, using Redis for DynamoDB is more complex than using DAX. Additionally, using Memcached for S3 content is inappropriate because Memcached is not designed to cache static files from S3, which is a job better handled by CloudFront.

Conclusion or Points to Memorize:

- DAX is the most effective solution for accelerating read performance in applications using DynamoDB, especially when dealing with high read request volumes for common data.
- CloudFront is optimal for serving static content from S3 due to its global network of edge locations that reduce latency and increase content delivery speed.
- Integrating specific AWS services based on their strengths and native compatibility ensures optimal performance and minimal implementation complexity.

Question 64:

An e-commerce company is looking for a solution with high availability, as it plans to migrate its flagship application to a fleet of Amazon Elastic Compute Cloud (Amazon EC2) instances. The solution should allow for content-based routing as part of the architecture. As a Solutions Architect, which of the following will you suggest for the company?

Correct Answer:

Use an Application Load Balancer for distributing traffic to the Amazon EC2 instances spread across different Availability Zones (AZs). Configure Auto Scaling group to mask any failure of an instance.

Explanation and Context:

The Application Load Balancer (ALB) is designed specifically for advanced load balancing of HTTP and HTTPS traffic, making it suitable for modern application architectures like microservices, which often require content-based routing. This feature of ALB allows it to route traffic based on the content of the request, such as path-based routing, which is beneficial for applications that deliver dynamic content to users. By distributing EC2 instances across different Availability Zones, the architecture ensures high availability and fault tolerance. The Auto Scaling group enhances this by ensuring that any instance failures are quickly compensated by automatically launching new instances to maintain the desired capacity, thus maintaining uninterrupted service.

Incorrect Options and Reasons:

1. **Use a Network Load Balancer for distributing traffic to the Amazon EC2 instances spread across different Availability Zones (AZs). Configure a Private IP address to mask any failure of an instance:** The Network Load Balancer is optimized for ultra-low latency and high throughput systems that operate at the connection level (Layer 4). It does not support content-based routing which is a requirement in the question.
2. **Use an Auto Scaling group for distributing traffic to the Amazon EC2 instances spread across different Availability Zones (AZs). Configure an elastic IP address (EIP) to mask any failure of an instance:** Auto Scaling groups do not distribute traffic; they are used for automatically adjusting the capacity of instances. While Elastic IPs can be used to mask the failure of individual instances by reassigning the IP, they do not aid in traffic distribution across multiple instances.
3. **Use an Auto Scaling group for distributing traffic to the Amazon EC2 instances spread across different Availability Zones (AZs). Configure a Public IP address to mask any failure of an instance:** Similar to the previous incorrect option, this setup misunderstands the role of Auto Scaling and public IP addresses. Auto Scaling is not for traffic distribution, and while public IPs can identify instances, they do not provide a mechanism for traffic distribution or high availability.

Conclusion or Points to Memorize:

- The Application Load Balancer is ideal for scenarios requiring content-based routing of HTTP/HTTPS traffic, facilitating advanced routing capabilities necessary for dynamic, user-centric content delivery.
- Availability is enhanced by using Auto Scaling groups in conjunction with ALBs across multiple Availability Zones, ensuring robust fault tolerance and continuous application availability.

• Understanding the distinct functionalities of different types of AWS load balancers (Application Load Balancer, Network Load Balancer, Classic Load Balancer) and other components like Auto Scaling groups and Elastic IPs is crucial for designing effective cloud architectures.

Question 65:

An IT company wants to review its security best-practices after an incident was reported where a new developer on the team was assigned full access to Amazon DynamoDB. The developer accidentally deleted a couple of tables from the production environment while building out a new feature. Which is the MOST effective way to address this issue so that such incidents do not recur?

Correct Answer:

Use permissions boundary to control the maximum permissions employees can grant to the IAM principals.

Explanation and Context:

A permissions boundary controls the maximum permissions that an IAM principal (user or role) can have. By defining permissions boundaries, you can delegate permissions management to developers or other employees without risking excessive permissions that could affect critical resources. Even if an employee assigns broad permissions to a role or user they manage, the effective permissions will be limited to the intersection of the permissions boundary and the attached policy. This prevents accidental or malicious high-level access to sensitive resources, like production databases, as was the case in the incident described.

Incorrect Options and Reasons:

1. **Only root user should have full database access in the organization:** Using the root user account for regular administrative tasks violates AWS best practices. The root user has unrestricted access to all resources and services in an AWS account, and its use should be limited to only a few account and service management tasks.
2. **The CTO should review the permissions for each new developer's IAM user so that such incidents don't recur:** While oversight is important, manually reviewing each new developer's permissions by high-level executives like the CTO is not scalable or error-proof. Automation and predefined policies like permissions boundaries are more effective and reliable.
3. **Remove full database access for all IAM users in the organization:** While limiting access is a good security practice, completely removing database access from all IAM users is impractical. Certain users will need this access to perform their roles effectively, particularly in database administration and maintenance tasks.

Conclusion or Points to Memorize:

- Implementing permissions boundaries is a crucial security control for managing the maximum allowable permissions assigned by IAM users, helping prevent excessive permissions that could lead to data loss or security breaches.
- Permissions boundaries provide a safeguard by ensuring that even if a policy grants broad permissions, the effective permissions are confined to what is defined in the boundary.

• It is important to design and apply permissions boundaries thoughtfully to balance security and usability without overly restricting necessary access for legitimate business and operational needs.